R E P O R T R E S U M E S

ED 019 974 LI 000 514

AN OPTICAL CHARACTER RECOGNITION RESEARCH AND DEMONSTRATION PROJECT.

LOS ANGELES COUNTY PUBLIC LIBRARY, CALIF.
COUNCIL OF LIBRARY RESOURCES INC., WASHINGTON,D.C.

PUB DATE 68

EDRS PRICE MF-$0.50 HC-$3.84 94P.

DESCRIPTORS- *OPTICAL SCANNERS, *LIBRARIES, *SYSTEMS DEVELOPMENT, COMPUTERS, INPUT OUTPUT, BOOK CATALOGS, COSTS, LIBRARY CIRCULATION, LIBRARY ACQUISITION, ELECTRONIC DATA PROCESSING, INFORMATION PROCESSING, LIBRARY TECHNICAL PROCESSES, EVALUATION, PERFORMANCE, EQUIPMENT EVALUATION, TEST RESULTS, OPTICAL CHARACTER RECOGNITION (OCR),

RESEARCH AND DEVELOPMENT OF PROTOTYPE LIBRARY SYSTEMS WHICH UTILIZE OPTICAL CHARACTER RECOGNITION INPUT HAS CENTERED AROUND OPTICAL PAGE READERS AND DOCUMENT READERS. THE STATE-OF-THE-ART OF BOTH THESE OPTICAL SCANNERS IS SUCH THAT BOTH ARE ACCEPTABLE FOR LIBRARY INPUT PREPARATION. A DEMONSTRATION PROJECT UTILIZING THE TWO TYPES OF READERS, SINCE THE CAPABILITIES OF EACH TYPE ARE SUITED TO DIFFERENT LIBRARY APPLICATIONS, WAS SET UP. LIBRARY REGISTRATION, BOOK CATALOG PRODUCTION, CARD CATALOG PRODUCTION, AND BOOK RECORD PRODUCTION WERE DEMONSTRATED ON THE PAGE READER AND LIBRARY CIRCULATION CONTROL ON THE DOCUMENT READER. FLOW CHARTS WERE DEVELOPED FOR EACH SUBSYSTEM AS WELL AS FORMATS FOR THE CONTROL INPUT FORMS AND THE PROCEDURES TO BE FOLLOWED. FOUR HUNDRED TWENTY-FOUR DOCUMENTS WERE TYPED USING A SPECIAL TYPE FACE AND WERE SCANNED BY THE PAGE READER, WHICH HAD AN ERROR RATE OF LESS THAN ONE PER CENT. THE MAGNETIC TAPE OBTAINED FROM THIS PROCEDURE WAS USED IN THE SECOND STEP IN WHICH A COMPUTER TRANSLATED AND TRANSFORMED THE DATA FOR INPUT INTO THE THIRD PROCEDURE, THE PRODUCTION OF THE CATALOGS OR REPORTS REQUIRED FROM THE ORIGINAL DOCUMENTS. THE CIRCULATION SYSTEM USED TWO-PART, INTERLEAVED TAB STOCK CARDS WHICH WERE IMPRINTED AT THE TIME OF CIRCULATION WITH THE NECESSARY INFORMATION. THE DOCUMENT READER READ THE CARDS AND PUNCHED THE INFORMATION ON THE SAME CARD WITH ABOUT A 25 PER CENT REJECT RATE. WITH PROPER CARE IN DATA PREPARATION AND CONVERSION AND WITH A DOCUMENT READER ADJUSTED FOR CORRECT PRINT INTENSITY, THE REJECTION RATE COULD HAVE BEEN LESS THAN 5 PER CENT. THE RECORDS WERE THEN TRANSFERRED TO MAGNETIC TAPE AND THE NECESSARY REPORTS MADE ON THE COMPUTER. A COST ANALYSIS WAS ALSO DONE, COMPARING THE FIVE DIFFERENT WAYS OF CONVERTING DATA INTO MACHINE-READABLE FORM. THIS STUDY SHOWED THAT OPTICAL SCANNING METHODS ARE LEAST EXPENSIVE ON A MONTHLY BASIS. IN GENERAL, DEMONSTRATIONS CONDUCTED DURING THE COURSE OF THE PROJECT INDICATED THAT WITH PROPER PREPARATION AND NORMAL ATTENTION TO DETAIL, THE ERROR AND REJECTION RATES OF BOTH PAGE AND DOCUMENT READERS COULD BE KEPT WELL WITHIN DESIRABLE LIMITS. THIS REPORT IS ALSO AVAILABLE FREE ON REQUEST AS LONG AS THE LIMITED SUPPLY LASTS FROM THE LOS ANGELES COUNTY LIBRARY, P.O. BOX 111, LOS ANGELES, CALIFORNIA 90053. (CM)

LL 000514

ERIC/CLIS
JUN 20'68
UNIVERSITY OF MINNESOTA

ED019974

AN OPTICAL CHARACTER RECOGNITION RESEARCH and DEMONSTRATION PROJECT

LOS ANGELES COUNTY

PUBLIC LIBRARY SYSTEM

LL 000514

Sponsored by the Council of Library Resources, Inc.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINION STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

AN OPTICAL CHARACTER RECOGNITION RESEARCH AND DEMONSTRATION PROJECT

prepared by

LOS ANGELES COUNTY PUBLIC LIBRARY
320 W. Temple Street
Los Angeles, California 90012

sponsored by

THE COUNCIL ON LIBRARY RESOURCES, INC.
Washington, D. C. 20036

PREFACE

Optical Character Recognition (OCR) is one of several input methods which can be utilized with computer systems. There are significant advantages and disadvantages inherent in using OCR in computerized library systems. Two types of readers, page readers capable of reading numerous typewritten or printed characters on many lines, or, document readers capable of reading a limited number of characters on one or two lines, will be of increasing use to libraries and other organizations in the future.

This study, which was undertaken by Los Angeles County Public Library with the support of the Council on Library Resources, Inc., represents research and development of prototype OCR - computer based library systems. Los Angeles County Public Library, which has been a leader in the development of automated book catalogs, is currently designing and implementing a computerized book catalog production system with optical character recognition as the input method.

ACKNOWLEDGMENTS

Los Angeles County Public Library acknowledges with gratitude the assistance of the numerous individuals, public and private organizations, and commercial concerns who provided information, material, support and encouragement. Particular thanks are due to the Council on Library Resources, Inc. (CLR), a major supporter of this project. The generous and enthusiastic cooperation of Mr. Verner W. Clapp, former president of CLR, and Mr. Laurence Heilprin, formerly of CLR, is most appreciated.

Los Angeles County Public Library personnel closely associated with the project included: Mr. Robert C. Goodwell, Assistant County Librarian, Mr. Ronald Zuckerman and Mr. Don Gill. Daniel, Mann, Johnson and Mendenhall of Los Angeles performed a major portion of the system design and development effort. The assistance of Dr. Robert Hayes, UCLA, during the course of this project was most appreciated.

Special thanks are due to the following commercial organizations for their excellent technical and informational services:

Addressograph-Multigraph Corp.
Control Data Corporation
Data Corporation
Diamond National Corporation
Farrington Manufacturing Company
Standard Register Company
International Business Machines
Philco Corporation
Recognition Equipment, Inc.
United California Bank
Atlantic Richfield Corporation

William S. Geller
County Librarian
June 1, 1968

TABLE OF CONTENTS

	Page
PREFACE	i
ACKNOWLEDGMENTS	ii
TABLE OF CONTENTS	iii
LIST OF EXHIBITS	iv
INTRODUCTION	1
SECTION I - Optical Character Readers	4
. Principles of Operation	5
. Hardware	8
. Operation	9
SECTION II - Library Input Requirements	17
. Requirements	19
. Analysis of Data Collection Costs	25
SECTION III - System Design	38
. Catalog/Book Record	39
. Registration Control	43
. Circulation Control	50
. Demonstration System	55
. Acquisition System	56
SECTION IV - Systems Demonstration	63
. Page Reader Demonstration	63
. Document Reader Demonstration	77
SECTION V - Project Summary	85

LIST OF EXHIBITS

TABLE	TITLE	PAGE
I	Optical Page Readers	10
II	Optical Document Readers	11
III	Journal Tape Readers	13
IV	Miscellaneous Readers	14
V	Page Reader Demonstration Error Summary	68

FIGURE		
1	Optional Book Charging Methods	21
2	Capabilities of Data Collection Equipment	22
3 (a)	Alternatives Worksheet	29
3 (b)	Formulas for Worksheet Computations	30
4	Computation for Case Study Worksheet	31
5	Factors Used in Sample Worksheet	34
6	Alternatives Worksheet for Case Study	37
7	Flow Chart-Card Catalog Program	44
8	Flow Chart-Book Catalog Program	46 - 47
9	Catalog Input Form	48
10	Registration Control Input Form	51
11	Flow Chart-Registration Control Program	52 - 53
12	Flow Chart-Circulation Control Program	58 - 59

LIST OF EXHIBITS (cont'd)

FIGURE	TITLE	PAGE
13	Catalog Production Input	65
14	Registration Control Input	66
15 (a)	Library Registration Report	70
15 (b)	Patron Registration Statistics	71
16	Book Catalog by Author	72
17 (a) & (b)	Card Printing	73 - 74
18	Book Catalog by Title	75
19	Book Catalog by Subject	76
20 (a)	Circulation Report	80
20 (b)	Circulation Statistics	81
21	Overdue Notices	82
22	Patron Registration Form	83
23	Circulation Transaction Form	83
24	Bibliographical Data Form	84

INTRODUCTION

The Los Angeles County Public Library is the fifth largest public library in the United States. Current support for the Library is approximately eleven million dollars per year. The Library maintains over 86 community outlets, services 18 County institutions, and provides bookmobile service to various outlying communities. The Library book collection exceeds 2.5 million volumes, with approximately 260,000 individual titles. Independent cities in Los Angeles County and unincorporated areas of the County comprise the area served by the Library. An ambitious building program is in progress with some 50 new library buildings contemplated in this decade. In addition, the Los Angeles County Public Library is a pioneer and leader in the development of book catalogs.

In 1965, the Los Angeles County Public Library received a research grant, CLR 346, from the Council on Library Resources, Inc., to investigate the uses of Optical Character Recognition (OCR) equipment as an input media for computerized library systems. This grant had the following objectives:

1. Produce a state-of-the-art report on optical scanning with emphasis on character recognition. This report, written by Ronald Zuckerman, was published in 1967. It was widely distributed and is now in short supply.

2. Develop and demonstrate the use of OCR in library applications. This required systems studies, selection of equipment for testing, the design of forms for both demonstrations and actual use, data preparation, and performance of OCR demonstrations.

3. Develop and test prototype systems which would utilize data files created through the use of OCR devices.

4. Report on the methods, tests performed and results of the research demonstration project.

This report outlines the design of such a research and demonstration program for the Los Angeles County Public Library. It represents a prototype system predicated upon the requirements and capabilities necessary for a data handling system which will satisfy the registration control subsystem, circulation control subsystem, the catalog production subsystem, and the acquisition control subsystem. In order to accomplish these objectives, engineering studies pertaining to the data rates, data formats, data system equipments and facilities, and other criteria have been performed. The results of these studies are presented in this report.

Section I of this report reviews the state-of-the-art of optical scanning, and presents a discussion of the principles, hardware, and operation of optical scanning systems.

Section II discusses the input requirements of a library, alternative methods of input and relative costs of these alternatives.

Section III describes the demonstration systems developed and includes Catalog Production, Circulation Control and Registration Control. An Acquisition System is conceptually designed, but was not developed nor used as part of the demonstration.

Section IV discusses the operation and results of the page reader and document reader demonstrations.

SECTION I

OPTICAL CHARACTER READERS

Optical scanning is a technique which permits the analysis of documents by machine, with the informational content digitalized for further machine processing. The information they contain may be characters, codes, marks or patterns. Documents may consist of such physical elements as pages, unit cards, journal tapes, or microfilm. Optical scanners may consist of devices which operate independently of other equipment; or, of devices which are interfaced to a computer system and operate under program control. The output from optical scanners may consist of punched cards, punched paper tape, magnetic tapes, or signals representing digitalized data.

This report will only discuss the following selected types of optical scanning techniques:

1. Optical Character Recognition (OCR), in which letters, numerals and symbols entered on documents are scanned. Characters may be handwritten, typed, printed or imprinted.

2. Optical Mark Reading (OMR), in which marks entered manually or by machine on documents are scanned.

3. Optical Code Recognition, in which codes representing numerals and other characters are entered on documents and subbsequently scanned.

4. Pattern Recognition, in which scanned images are digitally stored for recall and display.

PRINCIPLES OF OPERATION

The basic components of an optical scanning system are input documents prepared in accordance with scanning specifications, scanning hardware to process documents, and machine operation personnel.

Optical scanners recognize images by analysis of light absorption or reflection. The scanner accomplishes this by flooding the document with light, directing reflected light to a photomultiplier, converting light patterns captured in the photomultiplier to electrical energy (signals), analyzing these signal patterns against logic and memory circuitry, and, if a match with predetermined signals can be made, outputs machine-processable codes. Optical scanning hardware requires a document transport, a scanning mechanism, a logic mechanism, and an output mechanism. Optical scanners either scan a moving document once, with a fixed light; or, scan a stationary document one or more times with a moving light source.

It is predicted that optical scanning will become increasingly common in larger data processing installations during the 1970's. Scanners capable of reading hand printing and multiple fonts are becoming available at reasonable costs and standardization of font will allow even wider use of scanning for an increasing variety of documents. The present use of optical scanners is limited by economic considerations rather than by technological barriers.

In explaining the electro-mechanical aspects of this technique, it must be understood that the document transport moves the document from a feed station to a scan station, positions the document as necessary and moves the document to an output station. The speed of most optical scanning systems is limited by the document transport. Transport mechanisms range from simple two reel arrangements for handling paper journal tapes to complex mechanisms which will handle documents of various size, weight and format, which may be selectively stacked in several output hoppers. An important aspect of transport mechanisms is to keep documents within machine skew tolerances.

The function of the scanning mechanism is to view images and to convert the image into an electrical signal. Various approaches have been used to scan images. The major techniques which have been used or are in use today include:

1. Optical Masking
2. Flying Spot Scanning
3. Line Scanning
4. Mechanical Roster Scanning
5. Full Matrix Scanning
6. Electronic Tube Techniques

The logic mechanism of an optical scanner identifies the output from the scanning mechanism. This may be as simple as the determination of the existence or nonexistence of a mark, or as complex as identifying which of several thousand different characters is being viewed. Once the identification is made, coding into a computer-processable format becomes a straightforward matter.

There are two basic approaches to identifying scanned image output. These are the nearest match and exact match techniques. The first of these systems searches for the character definition closest to the optical output and identifies the two as the same. There are no rejects. In an exact match system, however, the image is defined with tolerance limitations in memory. If incoming images do not exactly match recorded images, a reject occurs. In practice, optical scanners usually combine a mixture of both techniques to keep rejects and misreads at low rates.

The function of the output mechanism is to convert the identified image into a computer-processable code which is entered into the computer, or on a storage medium, such as magnetic tape.

HARDWARE

Optical character readers range from machines capable of reading a single line of numeric characters is a single stylized font to machines which can read full pages of alphameric characters, special symbols and marks in a number of fonts. Certain machines are currently capable of reading a limited number of handwritten symbols when they are carefully inscribed within certain specifications

Basically, then, four types of optical character readers are now available. These include page readers, document readers, journal tape readers, and microfilm readers (currently film). The difference between page and document readers is the amount of data which can be read, and the speed of document throughput. Document readers are designed to handle a limited amount of data in restricted format (one or two lines) at high speed. Some document readers are capable of reading standard and mirror images of characters. Page readers operate at lower speed, but are generally capable of handling more complex arrangements of data with many lines per page.

Tables I to IV present some of the important characteristics of optical character readers which are or have been commercially available in the United States.

OPERATION

Documents which are to be optically scanned must meet specifications of the scanner and be compatible with the input requirements of the total data processing system. While there are certain similarities between equipments, each model currently available accepts, transports and reads only documents which meet rigid standards. In most cases, this means reading documents created for a specific machine.

Five principal factors are related to the design and encoding of scannable documents. These include document measurements, document layout, document stock, encoding medium, and encoding equipment. Optical scanners accept batched documents of a single size. Some scanners will accept only a standard size while others may be adjusted, within limits, to read documents of different dimensions. Journal tape readers may read paper tapes of a single width, or of several widths, but each width is a separate run.

The important aspects of document layout are margins, horizontal and vertical spacing of data, and field grouping. Margins or blank spaces are required on all four dimensions of any document or reel to be read

TABLE I OPTICAL PAGE READER

COMPANY	MODEL	CHARACTER SET	FONT(S)	PAGE SIZE (inches)	MAXIMUM SPEED* (Chars/Sec)	DOCUMENT FEED & TRANSPORT	SCANNER & LOGIC
Burroughs	Typed Page Reader	Alphameric Punctuation	Upper Case Elite	L:11 W:8.5	75	Vacuum, Drum	Flying Spot, Matrix Match
Control Data	915/Page Reader	Alphameric Punctuation Symbols	A.S.A. & other Upper Case	L:2.5-14 W:4-12	370	Vacuum, Conveyor Belt	Parallel Photocell, Matrix Match
Farrington	1P Page Reader	Alphameric Punctuation Symbols	Selfchek A.S.A. & IBM 1428	L:5.5-13.5 W:4.5-8.5	280	Vacuum, Drive Roller	Mechanical Disc, Stroke Analysis
	Selected Data Page Scanner	Alphameric Punctuation Symbols	Selfchek	L:5.5-13.5 W:4.5-8.5	200	Vacuum, Drive Roller	Mechanical Disc, Stroke Analysis
	3030 Series III	Alphameric Punctuation Symbols	Selfchek A.S.A.	L:5.6-13.5 W:4.5-8.5	440	Vacuum, Drive Roller	Mechanical Disc, Stroke Analysis
Philco	General Purpose Print Reader	Alphameric Punctuation Symbols	Multiple Fonts	L:3-11 W:5-8.5	2000	Vacuum, Conveyor Belt	Flying Spot, Matrix Match
Recognition Equipment	Electronic Retina Rapid Index Page Reader	Alphameric Punctuation Symbols	Multiple Fonts	L:3.25-14 W:4.88-14	2400	Vacuum, Drum & Conveyor Belt	Parallel Photocell, Matrix Match

*Reading speed in terms of chars./sec. should not be used for comparison purposes. Selective reading requirements, document handling speed, page layout, machine reject rates, etc., all affect actual production speed of OCR devices.

TABLE II OPTICAL DOCUMENT READER

COMPANY	MODEL	CHARACTER SET	FONT(S)	DOCUMENT SIZE (inches)	MAXIMUM SPEED (DOC/MIN)	DOCUMENT FEED & TRANSPORT	SCANNER & LOGIC
Farrington	1D Document Reader	Alphameric Punctuation Symbols	Selfchek, IBM 403 IBM 1428	L:2.2-6 W:2.75-8.5	440	Vacuum, Drive Roller	Mechanical Disc, Stroke Analysis
	Model 9SP	Numerals	Selfchek	Tab cards; 51 or 80 column	180	Vacuum, Drive Roller	Mechanical Disc, Stroke Analysis
	2S3C Self-Punch	Alphameric Punctuation Symbols	Selfchek	Tab cards; 51 or 80 column	550	Vacuum, Drive Roller	Mechanical Disc, Stroke Analysis
	Model 3010 Series 111	Alphameric Punctuation Symbols	Selfchek, IBM 1428 A.S.A.	L:2.25-8.5 W:2.0-6.0	440	Vacuum, Drive Roller	Mechanical Disc, Stroke Analysis
	Model 3020 Series 111	Alphameric Punctuation Symbols	Selfchek, IBM 1428E A.S.A.	Tab cards; 51 or 80 column	550	Vacuum, Drive Roller	Mechanical Disc, Stroke Analysis
General Electric	COC-5 Bar Font Reader	Numeric 1 Symbol	COC-5 E-13-B	L:4.75-9 W:2.5-4.1	1200	Vacuum Conveyor Belt	
IBM	1282	Numerals 3 Symbols	IBM 1428	Tab cards; 51 or 80 column	200	Friction, Clutch	Mechanical Disc, Matrix Matching
	1418 I, II	Numerals Symbols	IBM 407	L:5.875-8.75 W:2.75-3.67	420	Friction, Vacuum Drum & Conveyor Belt	Mechanical Disc, Matrix Matching

TABLE II OPTICAL DOCUMENT READERS (Cont.)

COMPANY	MODEL	CHARACTER SET	FONT(S)	DOCUMENT SIZE (inches)	MAXIMUM SPEED (DOC/MIN)	DOCUMENT FEED & TRANSPORT	SCANNER & LOGIC
IBM	1428 I, II, III	Alphameric Symbols	IBM 1428	L:2.25-4.25 W:3.25-8.75	400	Friction, Vacuum Drum & Conveyor Belt	Mechanical Disc, Matrix Matching
	1287 I, II	Numeric 5 letters	ASA, 1428 NOF, Gothic Handwritten	L:3-9 W:2.25-5.9	-		
Optical Scanning Corporation	opScan 288	Numeric Symbols 5 letters	A.S.A. Hand-written	L:4.5-8.5 W:2.5-4	600		
Philco	General Purpose Print Reader	Alphameric Punctuation Symbols	Multiple Fonts	L:3-11 W:5-8.5	360 (3x5 cards)	Vacuum, Conveyor Belt	Flying Spot, Matrix Matching
Rabinow	RUR3200-1	Numeric	Open Billing Type	L:5.5 W:8.63	400	Vacuum, Conveyor Belt	Parallel Photocells, Matrix Matching
RCA	5820 or 70/251	Numeric 5 Symbols	RCA N-2	L:2.5-4 W:2.5-8.5	1800	Vacuum, Conveyor Belt & Drum	Vidicon Scanner, Stroke Analysis
Recognition Equipment	Electronic Retina Document Carrier	Alphameric Punctuation Symbols	Multiple Fonts	L:2.5 W:5-8.5	1200	Vacuum, Conveyor Belt	Parallel Photocells, Matrix Matching

TABLE III JOURNAL TAPE READERS

COMPANY	MODEL	CHARACTER SET	FONT(S)	TAPES	MAXIMUM SPEED (LPM)
IBM	1285	Numeric Symbols	NCR, 1428	W:1.31-3.5 L:36-200	3000
	1287 II	Numeric 5 Alpha	ASA, NOF Gothic, 1428 Handwritten	W:1.31.4.5	-
NCR	420-2	Numeric Symbols	NCR	W:1.31-3.25 L:10-1200	3200

TABLE IV MISCELLANEOUS READERS

COMPANY	MODEL	COMMENTS
Bendix	Pattern Recog. Device	Currently being taught to classify Alpha characters
International Computers & Tabulators	P Q Monitor	Optical print quality monitor used in evaluation
Kidder Press Co.	081 Char. Tester	Used for enlarging images and testing print quality
Optical Scanning Corporation	Telescan	Reads 8 1/2 x 11 sheets onto magnetic tape for data phone transmission
Rabinow	4100-1	Alphameric tally roll reader
	4500	Alphameric (3 fonts) document reader

and minimum margins vary from machine to machine. This spacing of data, vertically and horizontally, must match the specifications of the scanner.

The grouping of data into fields is handled differently with various machines. Some machines separate the fields with actual characters or codes which the machine recognizes. Other machines recognize a field within specified x and y coordinates, while a combination of these techniques is incorporated into some scanners.

The characteristics of document stock also affect machine reading operations. Each manufacturer specifies or recommends a variety of stock characteristics The major document stock qualities which affect scanning are: color, weight or thickness, grain quality, finish and opacity (reflective qualities), and ink absorption characteristics. The finish and opacity of stock affects the way light reflects from a document, while paper which absorbs the ink unevenly distorts the images to be scanned. Imperfections such as creases, folds and dirt specks may also affect the scanning. Because of rigid scanning requirements, a number of manufacturers have developed specialized paper and other stock for use with specific scanners.

Documents to be scanned may be blank, preprinted in a single color, or preprinted in two or more colors prior to entry of data. Ink, carbon

paper impressions and pencil are the principal encoding mediums. The basic qualities of black inks, carbon, pencil, which affect scanning are blackness, reflectivity and density (splash characteristics).

The characteristics of machines used to encode documents also exert significant effect on scanning accuracy. On machines such as line printers, adding machines, and cash registers, the quality and condition of ribbons used is important. When carbon copies are to be scanned, the hardness of machine platens and precision of imprinters are factors which influence impression quality.

SECTION II

LIBRARY INPUT REQUIREMENTS

To a lesser or greater extent, all libraries perform similar identifiable functions. These functions which are exclusive of organizational framework may be classified as administration of the library, selection of materials, acquisition of materials, cataloging and/or indexing of materials, circulation of materials to borrowers, maintenance of materials and collections, reference and information services, collateral services, and facilities. The complexity and type of data which is used within these functions varies from library to library and between types of libraries (public, school, college, and research, industrial, etc.). Before discussing data used in these functions, it is necessary to further describe the activities performed within them.

Administration of the library includes the responsibility for planning, defining and/or determining policy, and reporting and carrying out the library program. In this capacity information and statistics may be collected concerning other tasks.

Selection of materials includes all activities related to the selection of materials.

Acquisition of materials includes the ordering of materials, the receipt of such materials, the maintenance of records related to orders, the production of book records such as shelf list cards, pockets, and spine labels, and preparation of statistical and other reports on operations.

Cataloging and/or indexing of materials provides access to and information about the library collection. To achieve this, catalogs, indexes and holdings lists must be produced.

Circulation of materials to borrowers includes the maintenance of records and control procedures concerning charges and discharges of library materials to authorized users, "holding" and "reserving" library materials, and the preparation of statistical and other reports on operations. Circulation includes borrower registration; charging and discharging of materials, and the maintenance of overdue, fine and delinquency procedures. Procedures must also be established concerning material loss or destruction, holds, reserves, the preparation of materials for shelving, and statistical reporting and information preparation.

Maintenance of materials and collections includes the inventory of materials, shelving and shelf-reading of collections, weeding of collections, rebinding of materials, maintenance of shelf list files, and preparation of statistical and other reports on operations.

Reference and information services include the answering of reference inquiries, inter-library loans, production of bibliographies, literature searches, control of use of noncirculating reference materials, and preparation of statistical and other reports on operations.

Collateral services and facilities widely vary between individual libraries. Typical services and facilities include conference rooms, user-typing facilities, study carrels, audio-visual facilities (not materials), and copying services.

REQUIREMENTS

In view of these functions, library data to be processed falls into four major categories: management data, circulation data, cataloging/indexing data, and selection/acquisition data. The latter is discussed in Section III.

Management Data

Management data usually have no special requirements in terms of character sets, print quality, or difficult input format problems. Therefore, practically any standard input equipment may be utilized for management data collection of various libraries.

Circulation Data

Input preparation for library circulation revolves about borrower registration and book charging. Information concerning borrower registration is straightforward, therefore, this input information need not require

any special characters or specialized data collection equipment. Book charging, while not requiring input of special characters, must be accomplished at a charging desk. Such input contains information about the borrower, his selection, and conditions of loan. Figure 1 depicts various ways in which automatic data processing equipment may be employed in circulation control.

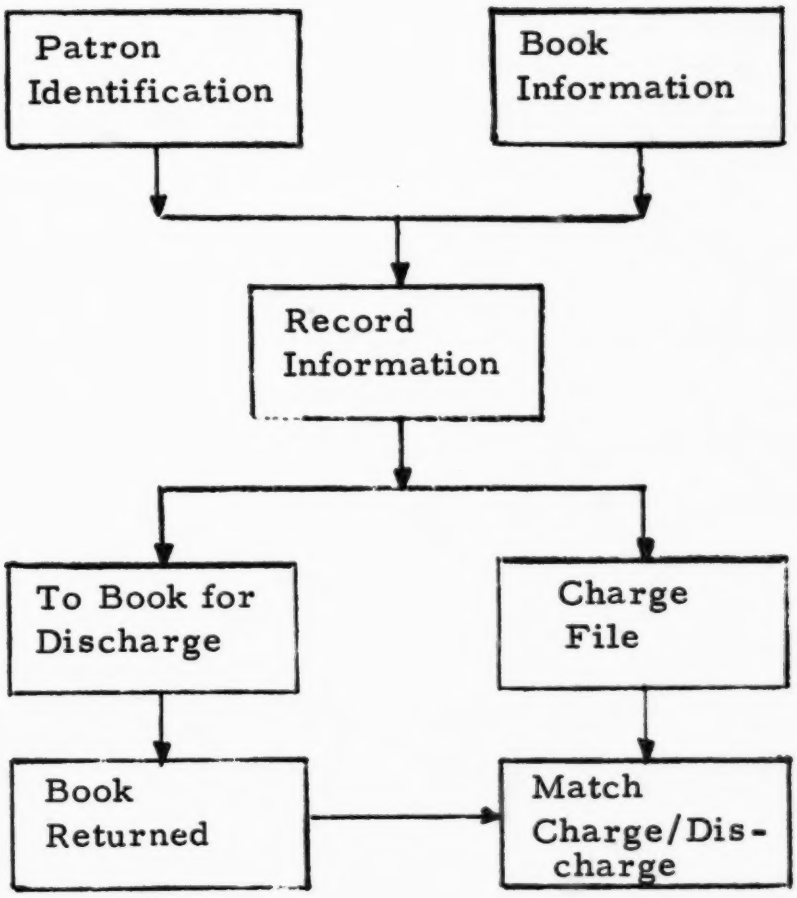
Cataloging/Indexing Data

Cataloging and indexing input preparation for libraries again represents wide variance between libraries. College and research libraries catalog materials in many languages, in great depth of detail and usually require extensive multi-lingual character sets. Only very specialized data collection equipment is capable of handling the input needs of such libraries. Public libraries without true research centers and school libraries deal less with multi-lingual materials, and often catalog in less depth. These less demanding catalogs are often restricted in their use of diacritical marks, and other alphabets, although upper and lower case alphabets are widely used. Figure 2 describes capabilities of currently available data collection equipment for such applications.

Input Preparation Methods

There are basically five different types of equipment available for preparation of input data:

BOOK CHARGING FUNCTIONS



Optional methods to accomplish book charging:

1. Patron identification recorded automatically (embossed card, etc.)
   Book information recorded manually.

2. Book information recorded automatically.
   Patron identification recorded manually.

3. Patron identification and book information recorded automatically.

4. Manual keying of patron identification and book information to produce records compatible with machine processing.

Figure 1 - Optional Book Charging Methods

CAPABILITIES OF DATA COLLECTION EQUIPMENT

DEVICE	Upper Case Alpha	Lower Case Alpha	0-13 Symbols	14-22 Symbols	Over 23 Symbols	Non-Roman Alphabets
Standard Printing Keypunches	x	x	x			
Standard Tape Perf. Type-writers	x	x		x		
Tape Perf. Type-writers equipped special Fonts	x	x				x
Standard Single Font OCR Page Readers	x	x		x		
Standard Multi-Font OCR Page Readers	x	x		x		
Special Purpose OCR Page Readers	x	x			x	x
Standard Magnetic Tape Writers	x	x		x		

Figure 2

1. Keypunches
2. Paper-tape punches
3. Magnetic tape writers
4. On-line terminals
5. Optical character readers

Magnetic ink character readers, pattern recognition equipment (such as mark sense, and bar codes), incremental measurement devices, and sound recording techniques are presently unsuitable for such use.

Keypunching is the oldest and most proven method of capturing data. It is also the most expensive. A maximum of eighty characters of information can be punched in each card. After a card is punched another operator "repunches" the same card on another machine, similar to a keypunch, called a verifier. With clean source data and experienced key punch operators and key verifiers an error factor of 3 percent is not uncommon. One way to approach 100 percent accuracy is to have the cards printed on a tabulator or small computer. Then clerks can visually check the source document against the machine printout, and if errors are found on the printout the corresponding error card is corrected.

Another high accuracy keypunching method is to have the cards printed (interpreted) as they are punched Again, clerks may be used to visually check the source data against the printed card. Error cards would be

repunched. It has been found that key punch and key verify operators can punch in a range from 60 to 1100 cards per hour, depending on the number of columns punched and the readability of the source data.

Punched paper tape is most frequently a secondary output from a cash register, add punch, or similar device. The tape is then fed into a computer, reformatted, audited and output on magnetic tape or punched cards. A machine printout is simultaneously performed permitting clerks to verify accuracy and totals. Corrections are keypunched and submitted to the computer with the error tape and a new corrected tape is written. When the output from paper tape results in cards, the error cards are replaced with corrected cards.

Magnetic tape writers are keyboard devices which write information directly on magnetic tape. A machine printout of the tape is necessary for visual verification. Where errors exist, correct cards are key punched and entered into the computer to produce an error free magnetic tape.

On-line terminals are used for keying information directly into a computer or peripheral processing system. Data is usually simultaneously typed or displayed on a CRT with on-line verification and correction procedures.

Optical character readers are capable of reading characters typewritten in a stylized font with great accuracy, but only when quality is

controlled This is the only method in which man and machine are able to "read" the same symbols. With character recognition, it is possible to typewrite the desired information, manually proofread the page, make corrections, additions and deletions, then optically read fully proved data, and output an error free magnetic tape.

ANALYSIS OF DATA COLLECTION COSTS

Determining the cost of converting library source data to machine readable form by alternative methods is a laborious task which must always be related to the constraints of the specific situation. It is futile to discuss the merits and costs of alternative data collection methods without relating to:

1. A specific application area, e.g., cataloging, ordering, etc.

2. A specific organization's application data capture requirements, e.g., character set, size of record, organization of records, etc.

3. The actual range of alternatives which may be considered by an organization. For instance, in-house data preparation may be mandatory, therefore service vendors could not be considered as an alternative.

5. The quantity of data to be converted in a specific time frame.

Rather than present general conclusions concerning the costs of optical character recognition as opposed to other library data conversion methods, DMJM has chosen to describe a hypothetical case study which will illustrate the costs of converting a large bibliographical data base to machine-processable form.

Case Study

As an example, a large public library system producing a book catalog by sequential camera desires to convert to another production method. Computerized book catalogs are to be considered as an alternative and a cost-benefit analysis must therefore be performed. Since a major cost in proceeding to a computerized catalog would be a huge one-time data conversion, alternative methods must be considered. Several restrictions to be placed on the considered alternatives are:

1. Data must be prepared entirely in the library.

2. Any equipment to be used in data conversion at the library must be rented rather than purchased.

3. The entire data conversion must be accomplished in 24 months.

4. Present library facilities must be utilized, therefore an actual computer or optical page reader cannot be placed on the premises.

The size of the data file to be converted is 300,000 titles/entries of 450 characters average. The records are 2,500 characters maximum and contain 20 possible types of variable length fields of which only 2 are mandatory (class number and title). Maximum field length is 999 characters and some fields, such as subject, may be repeated. The input character set to be used by the library consists of upper and lower case alpha, numeric and 18 symbols (punctuation, etc.). While the library publishes a foreign catalog, diacritical marks have been removed and only the Roman alphabet is used.

Alternatives – After considering the requirements, five basic alternatives for data conversion were considered:

1. Keypunching, prooflisting via service bureau, manual library proofing of lists, repunching of corrections, and service bureau conversion of cards to magnetic tape.

2. Typing/tape perforating, manual proofing of typed sheets, typing/tape perforating of corrections; and service bureau conversion of paper tape to magnetic tape.

3. Magnetic tape writing, prooflisting via service bureau, manual library proofing of lists, tape writing of corrections, and conversion to computer processable magnetic tape.

4. Typing on vendor supplied terminals, prooflisting at vendor service bureau, manual library proofing of lists, and retyping corrections on terminals.

5. Typing on electric typewriters with special fonts, manual proofing of typewritten sheets, retyping of corrections, and optical scanning of sheets at a service vendor (scanner to magnetic tape).

Procedures in Costing Alternatives – These would include:

1. Developing a worksheet (see Figure 3a).

2. Developing formulas for use in calculating items on worksheet. (See Figure 3b.)

3. Obtain unit costs for equipment, salaries, etc. (See Figure 5.)

4. Completing worksheets. (See Figure 4.)

Figure 6 represents a completed sample worksheet which analyzes costs of the five alternatives.

ITEM	COST/MONTH				
	Keypunch	Pap. Tape	Mag. Tape	On-Line	OCR
Personnel					
Keying					
Proofing					
Equipment					
Keying					
Service					
Computer Listing					
Media Conversion					
Space, Util. & Office Furniture					
Forms & Supplies					
Monthly Cost					
Cost for Conv. 300,000 Entries					

Figure 3a - Alternatives Worksheet

PERSONNEL COSTS

KEYING

1. (Entries/month x chars./entry) + 5% (errors rekeyed) = chars./month keyed

2. Char./month keyed ÷ keying speed x work hours x work days + keying personnel + 1/2 person (supervisor) required

3. (Keying personnel x cost/month) + diff. for supervisor = keying cost/month

PROOFING

4. (Entries x chars./entry) + 5% (errors reproofed) = chars./month proofed

5. Char./month proofed ÷ proofing speed x work hours x work days = proofing personnel + 1/2 person (supervisor) required

6. (Proofing personnel x cost/month) + diff. for supervisor = proofing cost/month

EQUIPMENT

7. No. keying personnel x cost/machine = equipment keying cost/month

SERVICE

8. Records/month x 1.05 = Records listed

9. Records listed x cost/record = listing cost

10. Entries/month x media conv. cost/entry = media conversion cost/month

SPACE, ETC.

11. Persons x cost/person = space cost/month

FORMS & SUPPLIES

12. $\frac{(\text{Forms/entry} \times \text{entries/month})}{1000} + 15\%) \times \text{cost/m} =$

forms cost/month

13. Persons x supply cost/person = supply cost/month

Figure 3b - Formulas For Worksheet Computations

1. (12,500 x 450) + 5% = 5,907,750 chars./month keyed or proofed

2. Keypunching or on-line typing cost

   5,907,750 ÷ 3600 x 6 x 20 = 13.5 + .5 operators

   (14 x $425) + $50 = $6,000.

3. Mag. or paper tape writing cost

   5,907,750 ÷ 3,850 x 6 x 20 = 12.5 + .5 operators

   (13 x $425) + $50 = $5,575.

4. Typing for OCR cost

   5,907,750 ÷ 4,250 x 6 x 20 = 11.5 + .5 operators

   (12 x $385) + $50 = $4,620.

5. Proofing cost for all methods

   5,907,750 chars./month proofed

   5,907,750 ÷ 13,500 x 6 x 20 = 3.5 + .5 proofers

   (4 x $385) + $50 = $1,590.

6. Keypunching equipment cost/month

   14 x $75 = $1,050.

7. Paper tape perforating cost/month

   13 x $75 = $975

8. On-line terminal cost/month

   14 x $1,150 = $16,100

9. Magnetic tape writing cost/month

   13 x $90 = $1,170.

Figure 4 - Computation for Case Study Worksheet

10. OCR typing cost

    12 x $25 = $300.

11. Computer listing for all methods except OCR

    13,125 x $.08 = $1,050.

12. Card to tape conversion

    12,500 x $.02 = $250

13. Mag. or Pap. tape to computer comp. mag. tape

    12,500 x $.05 = $725

14. OCR to Mag. tape conversion

    13,125 x $.22 = $2,887.50.

15. Space, etc.

    a. keypunch 18 x $20.50 = $372.

    b. paper tape 17 x $20.50 = $351.50

    c. mag. tape 17 x $20.50 = $351.50

    d. on-line 18 x $20.50 = $372.

    e. OCR 16 x $20.50 = $331.

16. Forms & supplies

    a. Tab cards

    $$\frac{(15 \times 12{,}500) + 15\%}{1{,}000} \times \$2.00 = \$432/\text{month}$$

    18 x $1.00 = $18.00/month

Figure 4 (cont'd)

b. Paper tape

$$\frac{(12,500) + 15\%}{1,000} \times \$12.00 = \$1,801/\text{month}$$

17 x $1.00 = $17.00/month

c. Mag. tape

$$\frac{(12,500) + 15\%}{1,000} \times \$30.00 = \$450/\text{month}$$

17 x $1.00 = $17.00

d. $$\frac{(12,500) + 15\%}{1,000} \times \$8.00 = \$120/\text{month}$$

18 x $1.00 = $18/month

e. OCR

$$\frac{(12,500) + 15\%}{1,000} \times \$8.00 = \$120/\text{month}$$

16 x $1.00 = $16.00/month

Figure 4 (cont'd)

ITEM	AV. COST/MONTH
Personnel	
Keypunch or machine operator	$425.00
Keypunch or machine op. supv.	485.00
Typist clerk	385.00
Typist clerk supervisor	435.00
Equipment (including maintenance)	
Keypunch	75.00
Paper tape typewriter	75.00
Magnetic tape writer	90.00
Electric Pinfeed typewriters	25.00
On-line keyboard + comp. time, line costs, etc.	1,150.00
Service	
Computer listing - 200 lpm upper/lower case printer	$.08/entry
Media conversion - 700 cpm to tape	.02/entry
Media conversion - mag. tape writer to computer tape	.05/entry
Media conversion - paper tape to mag. tape	.05/entry
OCR to magnetic tape	.22/entry

Figure 5 - Factors Used In Sample Worksheet

Space, Utilities	
Office furniture	$20.50/month/person
Forms & Supplies	
Tab. cards	$ 2.00/m
Typewriter sheets (8 1/2 x 11 pinfeed)	8.00/m
Perf. typewriter sheets + tape	12.00/m
Mag. tape	30.00/m entries
Miscellaneous supplies	1.00/month/person
Other Factors	
Typing speed (library data)	4,250 chars./hr.
Keypunching speed (library data)	3,600 chars./hr.
Tape writing or perforating	3,850 chars./hr.
On-line typing	3,600 chars./hr.
Proof Reading	13,500 chars./hr.
Average work hours/person/day	6
Average work days/month/person	20
Entries to be prepared/month	12,500
Re-keying - typing, etc. of errors found in proofing	5% of original number keystrokes
Characters/entry (average)	450
Supervisor timespent in data prep.	50%

Figure 5 (cont'd)

NOTES:

1. Personnel cost does not include fringe benefits.

2. Typewriter rental of $25.00/month is for IBM pin-feed selectrics.

3. Computer listing and card to tape conversion assumes an IBM 360/30 configuration with 1403N1 printer are average and setup is assumed in speeds indicated.

4. Space, utilities and office furniture cost/person/month was calculated as follows:

60 sq. ft./person @ .25/ft. =	$15.00
Utilities/person =	2.00
Office furniture (cost of chair, desk or file cabinet, etc., amortized over five years)	3.50
	$20.50

5. Miscellaneous supplies/person includes ribbons, writing tools, correction materials, etc.

Figure 5 (cont'd)

ITEM	Keypunch	Pap. Tape	Mag. Tape	On-line	OCR
Personnel					
Keying	$ 6,000	$ 5,575	$ 5,575	$ 6,000	$ 4,620
Proofing	1,590	1,590	1,590	1,590	1,590
Equipment					
Keying	1,050	975	1,170	16,100	300
Service					
Computer Listing	1,050	1,050	1,050	1,050	-
Media Conversion	250	725	725	-	2,887
Space, Utilities and Office Furniture	372	351	351	372	331
Forms & supplies	450	197	467	138	136
Monthly Cost	$10,762	$10,463	$10,928	$25,250	$ 9,864
Cost for conv. 300,000 entries	$258,288	$251,112	$262,272	$606,000	$236,736

This assumes all methods could utilize the same original source data (3x5 card). In practice, an input worksheet might be necessary for all except the OCR method. This could amount to a sizeable cost. Also, it is possible that machine operators are not as readily available as typists.

Figure 6 - Alternatives Worksheet

SECTION III

SYSTEM DESIGN

The Los Angeles County Public Library's data system is designed to provide maximum reliability and flexibility of operations. The design includes the capability to optically read typewritten documents and batch process these data.

The data facilities and equipment configurations in this system design were used to determine the feasibility of an OCR approach toward library data processing. The system was designed to determine realistic error rates and difficulties which might be encountered in the operation of a typical system. The design is based upon relatively simple and logical operating procedures, thereby demonstrating a practical working EDP oriented system for the Los Angeles County Public Library.

A framework for the progressive incorporation of present and future technological advances which will extend the total capability to convert optically read source data into processed reports has been developed. The structure of the system, while permitting sequential implementation, was not planned within a total systems approach for any individual library.

The subsystems designed for the prototype demonstrations were catalog and book records production, registration control, and circulation control. An acquisition system is discussed but was not designed nor demonstrated. These subsystems are described later in this section of the report.

It was decided to demonstrate two types of OCR devices, a page reader and a document reader, since the unique capabilities of each type of machine are suited to different library applications. Since no data processing center or installation in the Southern California area was known to have both capabilities, two separate demonstrations were planned. Library registration, book catalog production, card catalog production, and book record production would be approximated in the page reader demonstration. Library circulation control would be approximated in the document reader demonstration.

CATALOG/BOOK RECORD

There are two basic types of library catalogs; card catalogs and book catalogs. In card catalogs a single unit card is produced in a specified format. Headings are then added to unit cards providing access to a single work in several ways. For example, a unit card will usually list the author of a work first. The title and subject of the work may be added to copies of the unit card. When separately filed, these three cards provide

three access points to the work for the user. Book catalogs generally follow similar record formats.

Book records refer to shelf list cards, book cards, spine and pocket labels. These items all contain portions of bibliographical data which may be captured in the acquisition of materials or cataloging process.

Mechanized book catalog procedures all employ some manual procedures. The balance between manual and machine operations varies with different systems. A computer may be used in a limited capacity, that is as a printer for listing decks of punched cards which have been manually sequenced. In a highly computerized book catalog system, a complete bibliographical record may be prepared once in a machine-processable form. Subsequent computer processing might include extracting data from master records for catalog entries, formatting entry data, machine sequencing or alphabetizing of entries in a dictionary catalog or individual sections of a divided catalog, deletion of repetitious heading information, maintenance of master and other files, and creation of multi-columned catalog pages which include fixed-spaced or proportionally spaced characters and justified or unjustified right hand margins in columns.

The minimal hardware capabilities and software requirements for computerized book catalog production increase with the number and/or type of operations to be performed on the computer. It is a most difficult

matter to accurately estimate system design and programming costs for a catalog system of this complexity.

To machine file catalog data in accordance with ALA or similar filing rules presents formidable problems. Computers sort fixed length fields of data strictly character by character. A typical collating sequence or ascending character order for computer filing is space, A-Z, and 0-9. A period (.) might file after the space and before the character A. Non-standard treatment of characters of data to be sorted must be programmed into the computer. For example, if U. S. is to be filed as United States, a special subroutine must be written to handle such cases. To develop a workable machine-filing system requires compromise. Data will require rearrangement in the input-preparation stage, and special symbols with programmed filing uses must be added to records. Filing rules will need to be modified on occasion for the benefit of the computer, and, quite possibly, the catalog user. Machine filing is discussed in some depth in a work issued by the Bro-Dart Foundation: Hines, Theodore C., and Harris, Jessica L. Computer Filing of Index, Bibliographic and Catalog Entries, 1966.

A single source record concept is difficult to develop for bibliographical data. Within catalog records are units of data such as author, title, notes, which may be designed as fields. Fields may be referenced by

tags. For example, the class number field might be identified by the tag 1; the author field by the tag 2; etc. Unfortunately, fields may contain extremely variable length data. Any given catalog record may contain data for only a selected number of fields. There is even lack of agreement in the library world as to how fine a division of data should be made when separating bibliographical data into fields. Library of Congress, with Project MARC, has completed considerable work dealing with organization of input data.

Other factors which affect computer catalog systems are master and supplement catalog frequency of issue, the dictionary arrangement or divided nature of the catalog, readability and format requirements for printed copy, and entry density on the printed page.

Demonstration System

DMJM developed a system to demonstrate the approximation of the three types of book catalog production, by author sequence, by title sequence and by subject sequence. The similar production of card catalogs and book records was also designed as part of the system.

In designing the system it was agreed that the sorting and ordering of the input data would be done prior to entry of the data into the system. These functions have been successfully mechanized and duplication of this effort would not add to the findings of this research project.

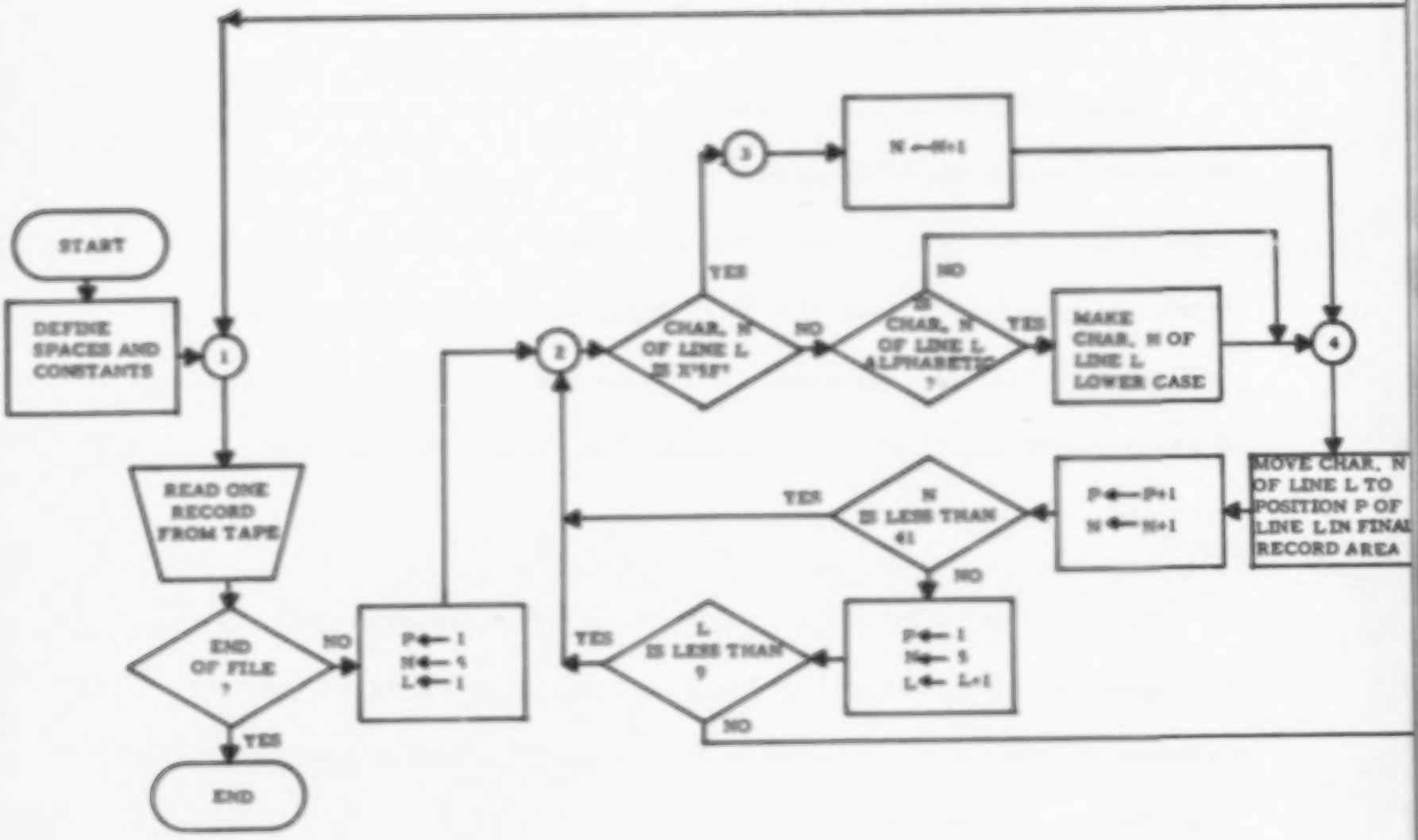
There are three procedures in the system. The first consists of preparing the input forms for the page reader and processing these forms through the reader. A magnetic tape is obtained for input into the second procedure which translates and reformats the data for input into the third procedure. The final procedure is the production of actual catalogs.

These procedures were necessary for the demonstration, although in an actual system the reformatting of the data would probably not be necessary. The Los Angeles County Public Library had already purchased typewriter equipment with ASA font, Rabinow CDC 915 font, and other OCR fonts. The demonstration was performed on a CDC 915 page reader. A printer with upper and lower case printing capability was required for the production of the catalogs.

The flow charts for the computer programs to prepare the book and card catalogs are presented in Figure 7. The flow charts for the book records program are presented in Figure 8. The input format for the page reader is presented in Figure 9.

REGISTRATION CONTROL

Registration control consists of providing a means of identification and the maintenance of information files concerning patrons. A master file would contain a registration number, the patron's name, address, and telephone number, a code designating type of patron, such as male


START
DEFINE SPACES AND CONSTANTS
1
READ ONE RECORD FROM TAPE
END OF FILE ?
NO
YES
END
P ← 1
N ← 1
L ← 1
2
CHAR. N OF LINE L IS X'5F'
YES
3
N ← 65+1
NO
IS CHAR. N OF LINE L ALPHABETIC ?
YES
MAKE CHAR. N OF LINE L LOWER CASE
4
MOVE CHAR. N OF LINE L TO POSITION P OF LINE L IN FINAL RECORD AREA
P ← P+1
N ← N+1
N IS LESS THAN 41
YES
NO
P ← 1
N ← 1
L ← L+1
L IS LESS THAN 9
YES
NO

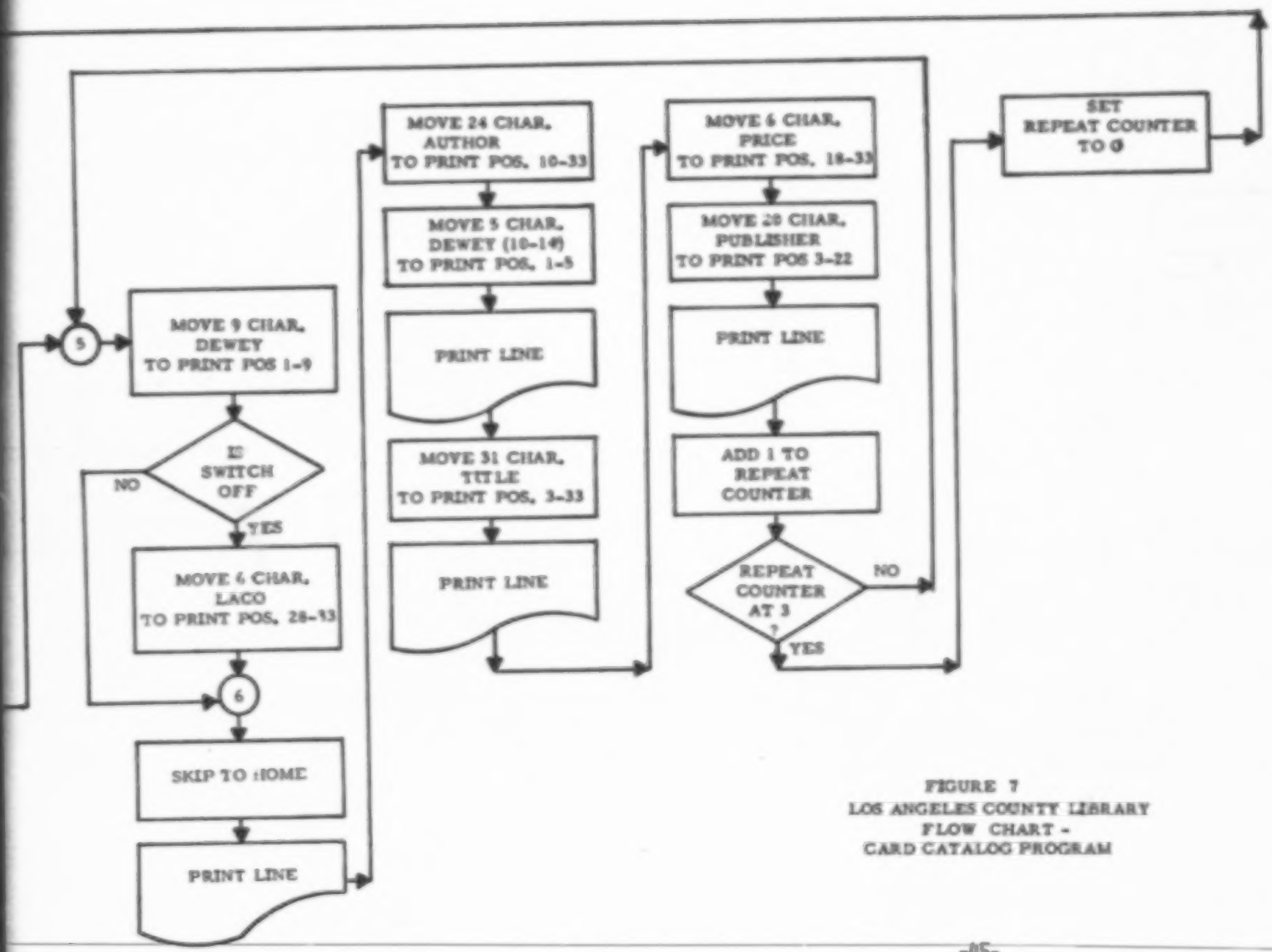
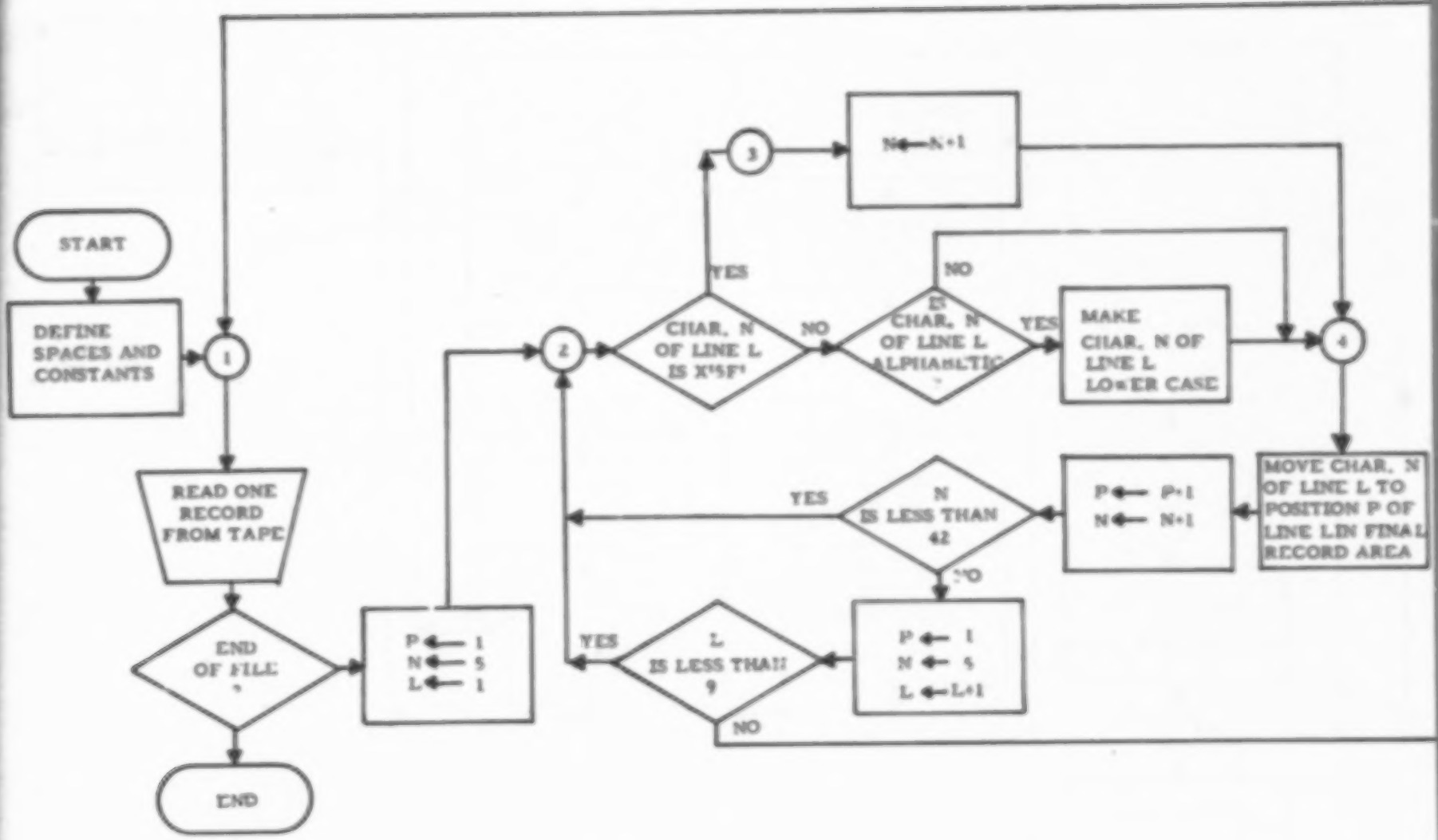


FIGURE 7
LOS ANGELES COUNTY LIBRARY
FLOW CHART -
CARD CATALOG PROGRAM


START
DEFINE SPACES AND CONSTANTS
1
READ ONE RECORD FROM TAPE
END OF FILE ?
P ← 1
N ← 5
L ← 1
END
2
CHAR. N OF LINE L IS X'5F'
YES
3
N ← N+1
NO
IS CHAR. N OF LINE L ALPHAMETIC
YES
MAKE CHAR. N OF LINE L LOWER CASE
4
MOVE CHAR. N OF LINE L TO POSITION P OF LINE LIN FINAL RECORD AREA
P ← P+1
N ← N+1
N IS LESS THAN 42
YES
NO
P ← 1
N ← 5
L ← L+1
L IS LESS THAN 9
YES
NO

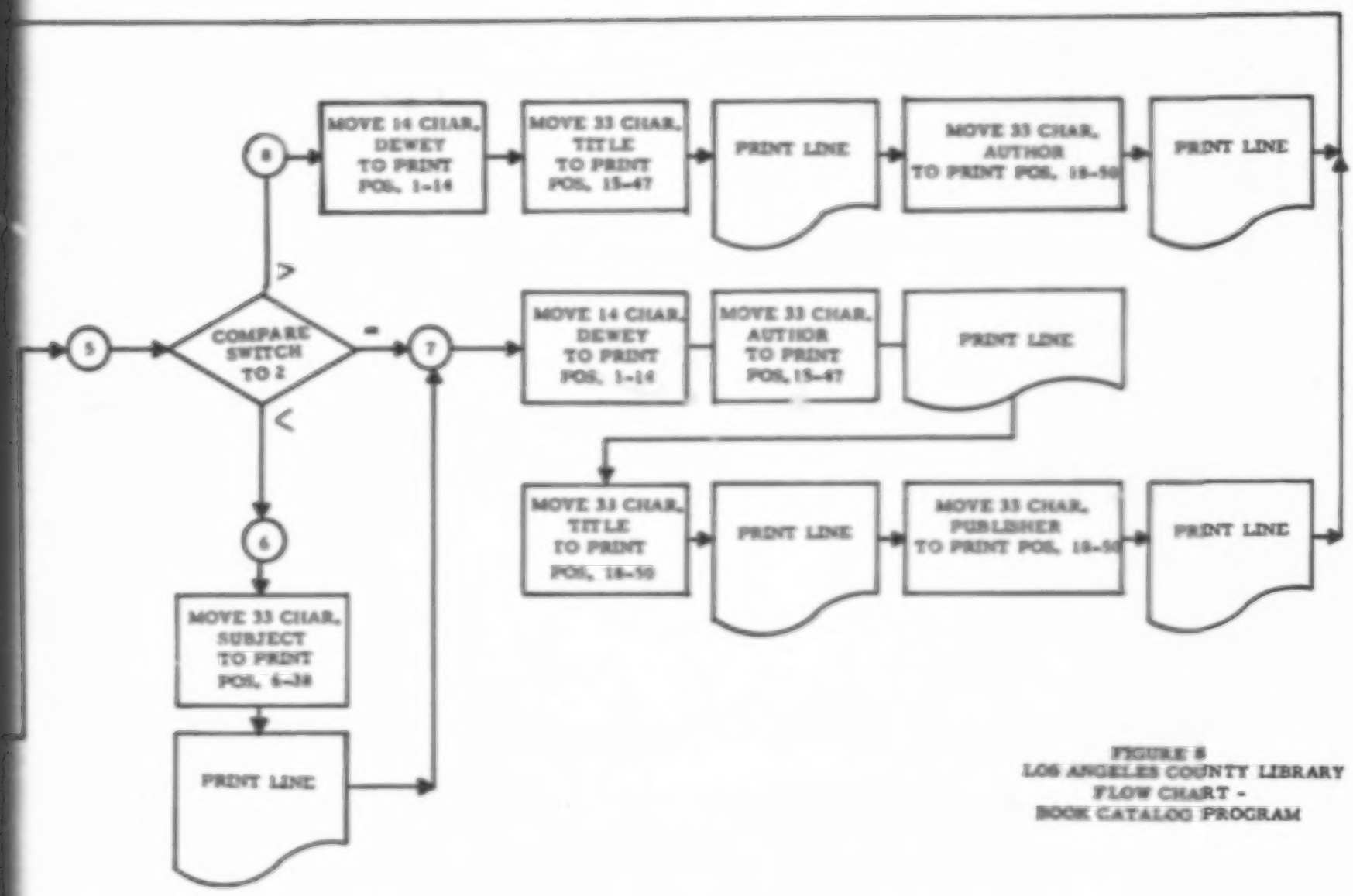

8
MOVE 14 CHAR. DEWEY TO PRINT POS. 1-14
MOVE 33 CHAR. TITLE TO PRINT POS. 15-47
PRINT LINE
MOVE 33 CHAR. AUTHOR TO PRINT POS. 18-50
PRINT LINE
>
5
COMPARE SWITCH TO 2
=
7
MOVE 14 CHAR. DEWEY TO PRINT POS. 1-14
MOVE 33 CHAR. AUTHOR TO PRINT POS. 15-47
PRINT LINE
<
6
MOVE 33 CHAR. SUBJECT TO PRINT POS. 6-38
PRINT LINE
MOVE 33 CHAR. TITLE TO PRINT POS. 18-50
PRINT LINE
MOVE 33 CHAR. PUBLISHER TO PRINT POS. 18-50
PRINT LINE

FIGURE 8
LOS ANGELES COUNTY LIBRARY
FLOW CHART -
BOOK CATALOG PROGRAM

TAG	DATA (ALL ALPHAMERIC)	MAX. CHARS. (INCLUDES DESIGNATORS)
0	CLASS OR DEWEY NUMBER	33
1	AUTHOR	33
2	TITLE	33
4	PUBLISHER AND DATE	33
7	ANNOTATION	33
8	SUBJECT	33
K	PRICE	33
L	COPY NUMBER (LACO____)	33
8 LINES TOTAL		274 x 2 FOR UPPER-LOWER CASE FUNCTIONS 548 POSITIONS OF MASTER RECORD TO BE BUILT

NOTES: TAGS WILL ALWAYS BE IN SEQUENCE

READING LINE WILL ALWAYS BE 40 CHARACTERS.

Δ WILL BE USED AS AN UPPER CASE DESIGNATOR.

DESIGNATORS WILL BE USED BEFORE EACH SUBJECT CHARACTER.

Figure 9 - Catalog Input Form

or female, student or non-student, married or single, and any other distinction required for statistical purposes. All transactions against the master file would be processed on a periodic basis. These transactions would include such items as change of name, change of address, or change of the patron code. A new master file would be produced each time the program was run.

In an operating system it might be desirable to maintain two files, one ordered by registration number and another ordered by patron name. All requests for new cards would be processed against the name ordered file which would prevent a patron from requesting a new card when he was delinquent or had lost his old card.

The basic input into the system would be specially prepared documents to be read by optical readers.

Demonstration System

It was decided to demonstrate the registration control system on an optical page reader. Even though the County Library designed a registration document to be read by optical reader, it was decided to use the same format for the registration control demonstration as used in the book catalog demonstration. This eliminated the writing of an additional program.

A master file and a transaction file would be prepared for the demonstration. Both files would be ordered by the patron name. The transactions would consist of code changes, change of address, new patron addition, renewing cards, replacing lost cards, and patron delinquency.

The output consists of a listing of the updated file (a new file is not created on cards or tape) and a summary of the number of patrons in each code.

Figure 10 presents the format of the input document to the page reader. Figure 11 is a flow chart of the computer program.

CIRCULATION CONTROL

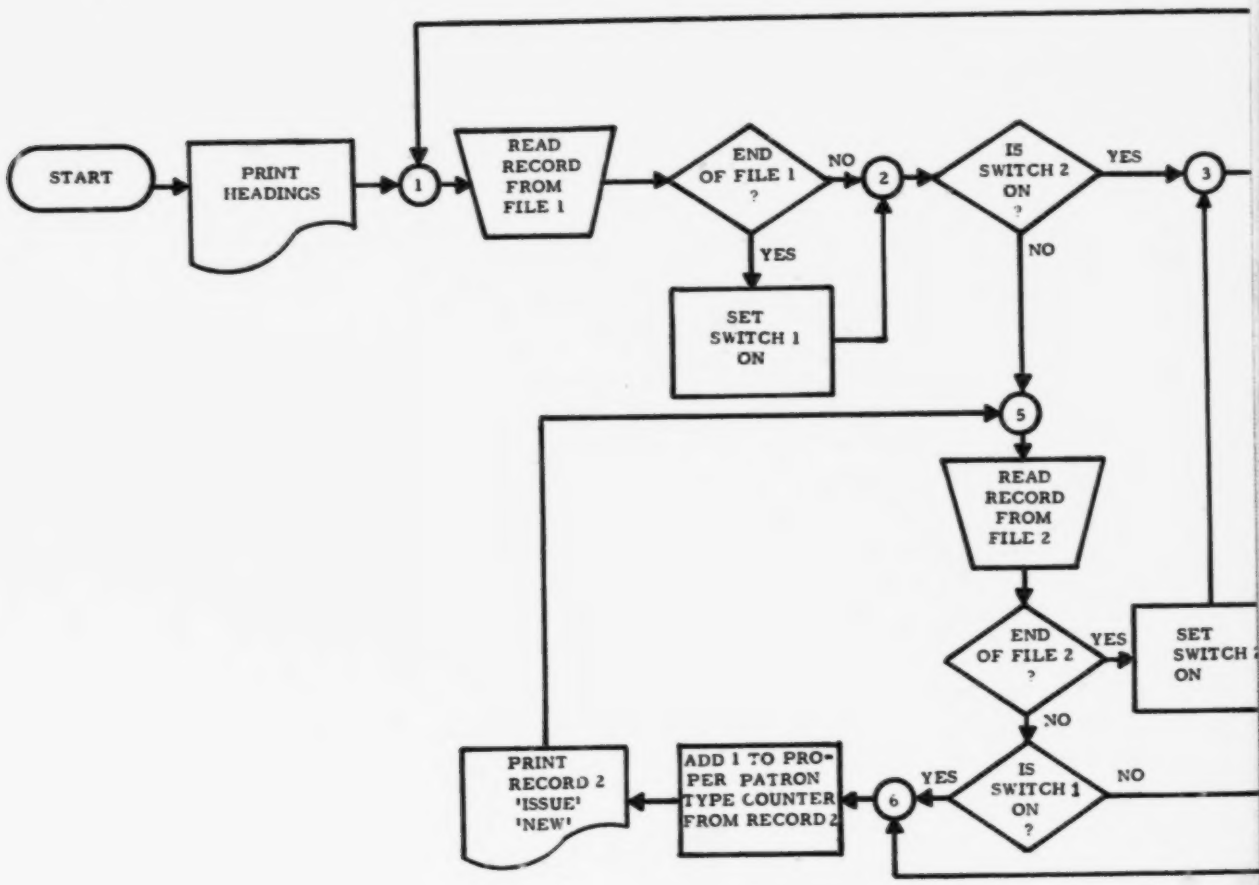
Circulation control functions include the charging of library materials, the discharing of library materials, the preparation of overdue notices, and fining and delinquency accounting procedures.

The charging of materials is the identification of the patron by name or number, identifying the material charged, and conditions pertaining to the transaction such as date charged and loan period.

The discharging of materials is the purging of the charge files by the return of the materials, and/or payment for the loss of the materials. Overdue notice preparation, and the fine and delinquency procedures are initiated during the purging of the charge files.

TAG	DATA	MAX. CHARS.
1	PATRON NUMBER/LIBRARY NUMBER	7 & 3
2	PATRON NAME	35
3	PRIOR PATRON NAME (if any)	35
4	ADDRESS OF PATRON (if change of address)	35
5	PRIOR ADDRESS (if change of address)	35
6	PRESENT CITY & ZIP CODE	21 & 5
7	TELEPHONE	7
8	CODE/NEW CARD/LOST CARD/RENEWAL/ NAME CHANGE/ADD DELINQUENT LIST/ REMOVE DELINQUENT LIST	1 each (7)
8 lines total		320 positions of record to be built (40 character line for scanning).

Figure 10 - Registration Control Input Form


START
PRINT HEADINGS
1
READ RECORD FROM FILE 1
END OF FILE 1 ?
NO
2
IS SWITCH 2 ON ?
YES
3
YES
SET SWITCH 1 ON
NO
5
READ RECORD FROM FILE 2
END OF FILE 2 ?
YES
SET SWITCH 2 ON
NO
IS SWITCH 1 ON ?
YES
NO
6
ADD 1 TO PROPER PATRON TYPE COUNTER FROM RECORD 2
PRINT RECORD 2 'ISSUE' 'NEW'

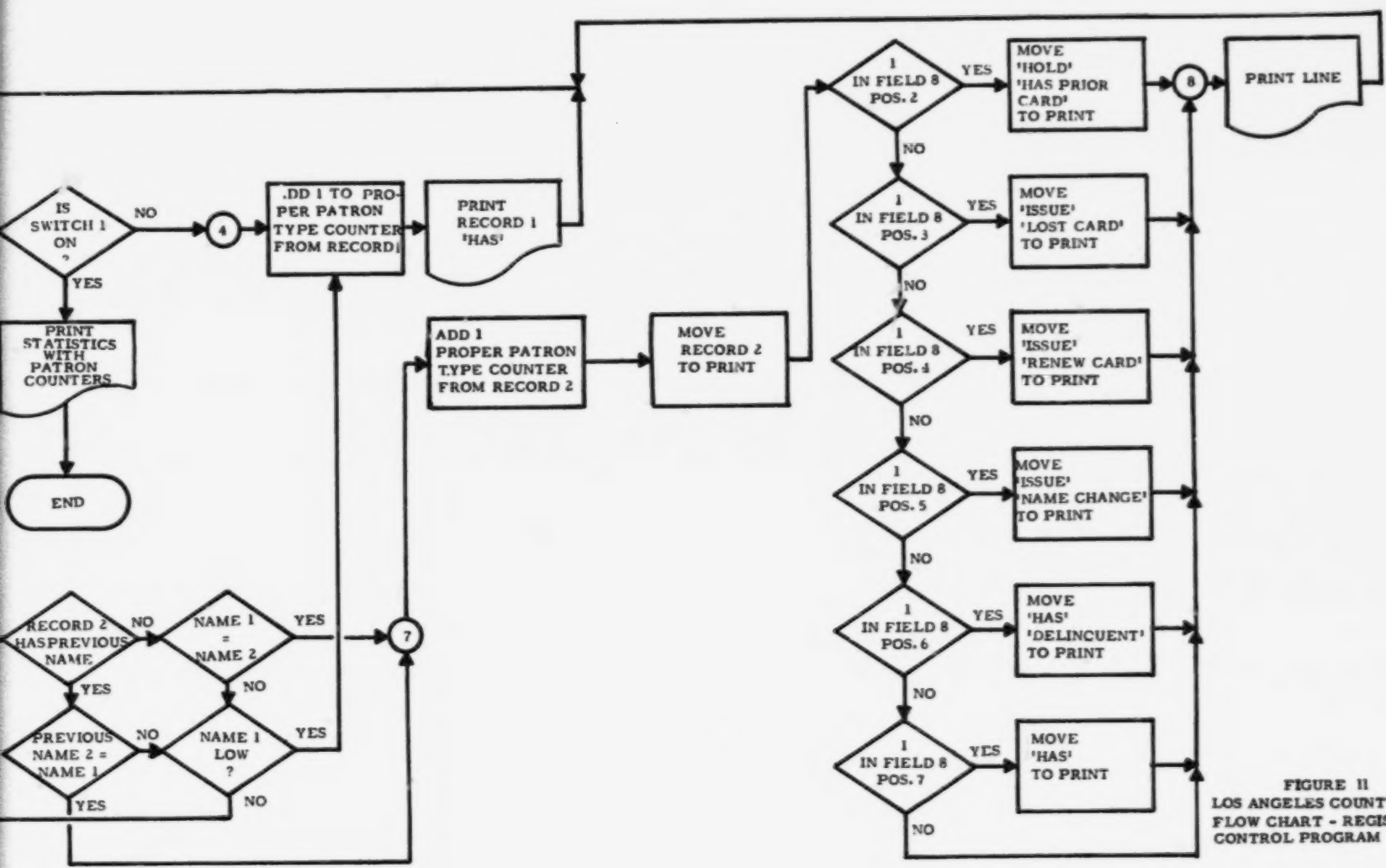

IS SWITCH 1 ON ?
NO
YES
4
ADD 1 TO PROPER PATRON TYPE COUNTER FROM RECORD 1
PRINT RECORD 1 'HAS'
PRINT STATISTICS WITH PATRON COUNTERS
END
ADD 1 PROPER PATRON TYPE COUNTER FROM RECORD 2
MOVE RECORD 2 TO PRINT
RECORD 2 HAS PREVIOUS NAME
NAME 1 = NAME 2
PREVIOUS NAME 2 = NAME 1
NAME 1 LOW ?
7
1 IN FIELD 8 POS. 2
MOVE 'HOLD' 'HAS PRIOR CARD' TO PRINT
8
PRINT LINE
1 IN FIELD 8 POS. 3
MOVE 'ISSUE' 'LOST CARD' TO PRINT
1 IN FIELD 8 POS. 4
MOVE 'ISSUE' 'RENEW CARD' TO PRINT
1 IN FIELD 8 POS. 5
MOVE 'ISSUE' 'NAME CHANGE' TO PRINT
1 IN FIELD 8 POS. 6
MOVE 'HAS' 'DELINCUENT' TO PRINT
1 IN FIELD 8 POS. 7
MOVE 'HAS' TO PRINT

FIGURE 11
LOS ANGELES COUNTY LIBRARY
FLOW CHART - REGISTRATION
CONTROL PROGRAM

An ultimate system would be a real time control of the charging and discharging of materials. This would involve remote terminals capable of reading patron and material identification, connected to a central processor and a large random access storage.

The status of the patron, whether in good standing or delinquent, could thus be determined immediately. It could also prevent the unauthorized use of a reported lost card. In discharging of materials any overdue fines would be determined by the computer and output at the remote terminal.

A non real time system can be designed to perform some of the same functions but the immediate control would be missing. This system would require some method of recording the charging information for later input into an automated system. Two approaches to charging may be taken in developing an automated circulation system: information may be recorded via transaction devices such a machine-readable badges, or cards; or information may be manually keyed or written in mark-sense codes. The use of manual keying or writing to capture charges introduces an unacceptable potential for error in an application where pressure and volume are high. If charging information is to be taken from plates or cards, embossed or debossed plates may be used to imprint data which is subsequently optically scanned. The ordering or creation of a file of charges completes the charging operation.

DEMONSTRATION SYSTEM

Three files were developed for the demonstration. A group of charge documents were read on an optical document reader to prepare the charge file. A subset of these documents was used to prepare a discharge file. A patron file was also simulated. The charge and discharge files were ordered on book number. The information in the two files consisted only of the book number and patron number.

The charge file was compared to the discharge file. On a match the Dewey number was extracted from the patron file and used in the determination of the statistical report. If there was no match, four conditions could occur:

1. The book was in normal circulation and its number and patron number were printed.

2. The book was due at this time. The book number, patron's number, and name and address were printed from the patron file.

3. The book was overdue. In such circumstances the printout is the same as for a book due, but, in addition, a card is punched to prepare overdue notices.

4. The book is far delinquent. The printout is the same but with the addition of the far delinquent column which is marked.

A statistical report is prepared on the results of the circulation control program. These results include the total number of books charged, returned, in circulation and the total number in circulation by Dewey number.

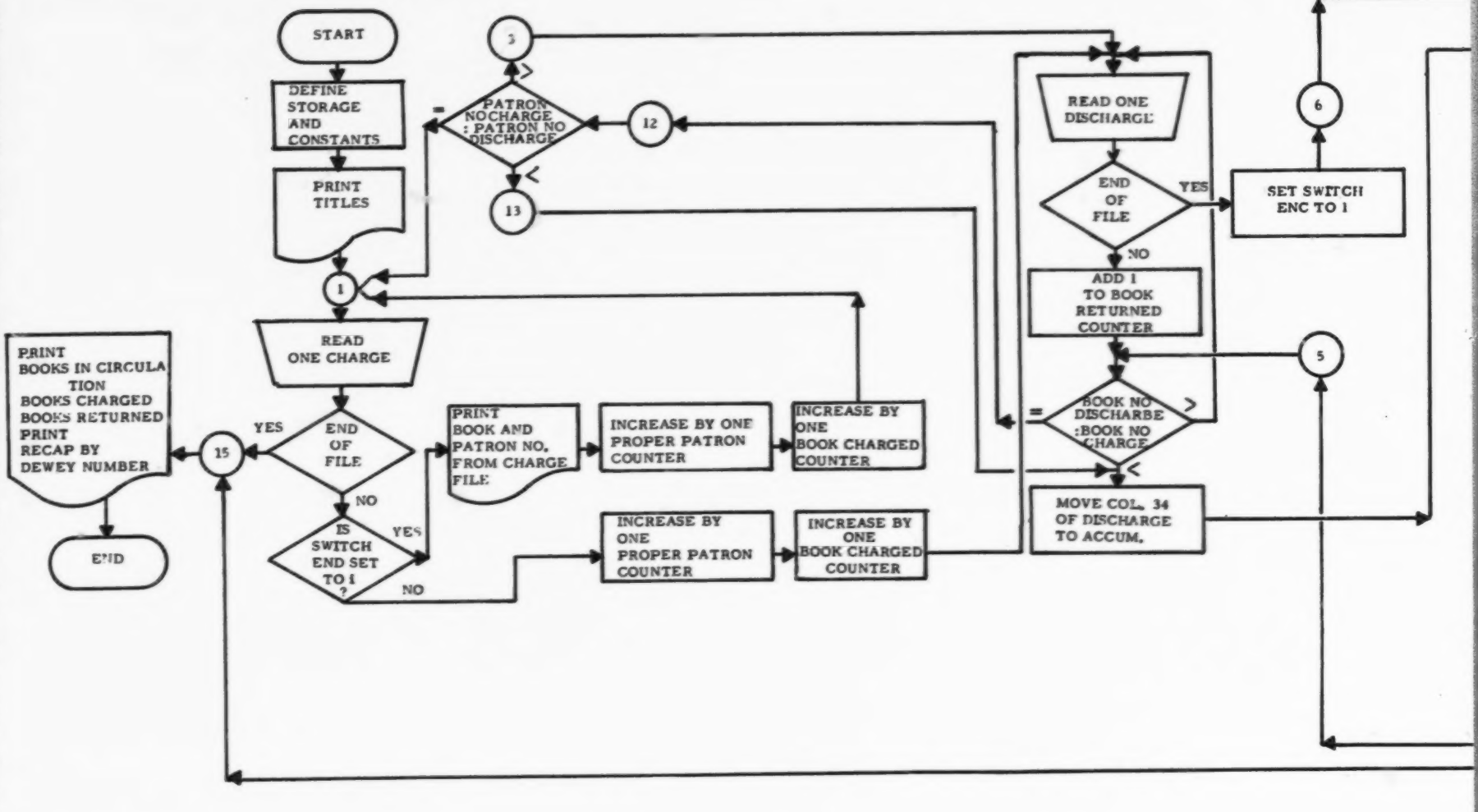
The card outputs for overdue books were used to prepare overdue notices.

Figure 12 is a flow chart for the circulation control program.

ACQUISITION SYSTEM

The Los Angeles County Public Library performs centralized acquisitions for the entire library system. This entails book evaluation, ordering functions, fund accounting and receiving activities. Initially, materials are reviewed for addition to the county collection by central library staff. Materials evaluated are presented at order meeting with order forms (3 x 5 cards) for regional librarians. Regional librarians determine the number of copies of each book required and submit the order form, retaining a copy for region use. The central staff prepares a multiple copy order form with the total number of copies requested.

After the multiple copy orders are prepared, the orders are sorted by publisher and/or vendor and converted to order lists (full size pages) containing many works. These lists are priced out against discounts if


START
DEFINE STORAGE AND CONSTANTS
PRINT TITLES
1
READ ONE CHARGE
YES
END OF FILE
NO
IS SWITCH END SET TO 1 ?
YES
NO
PRINT BOOK AND PATRON NO. FROM CHARGE FILE
INCREASE BY ONE PROPER PATRON COUNTER
INCREASE BY ONE BOOK CHARGED COUNTER
INCREASE BY ONE PROPER PATRON COUNTER
INCREASE BY ONE BOOK CHARGED COUNTER
INCREASE BY ONE PROPER PATRON COUNTER
INCREASE BY ONE BOOK CHARGED COUNTER
15
PRINT BOOKS IN CIRCULATION BOOKS CHARGED BOOKS RETURNED PRINT RECAP BY DEWEY NUMBER
END
3
PATRON NOCHARGE : PATRON NO DISCHARGE
12
13
READ ONE DISCHARGE
END OF FILE
YES
NO
SET SWITCH ENC TO 1
6
ADD 1 TO BOOK RETURNED COUNTER
5
BOOK NO DISCHARBE :BOOK NO CHARGE
MOVE COL. 34 OF DISCHARGE TO ACCUM.

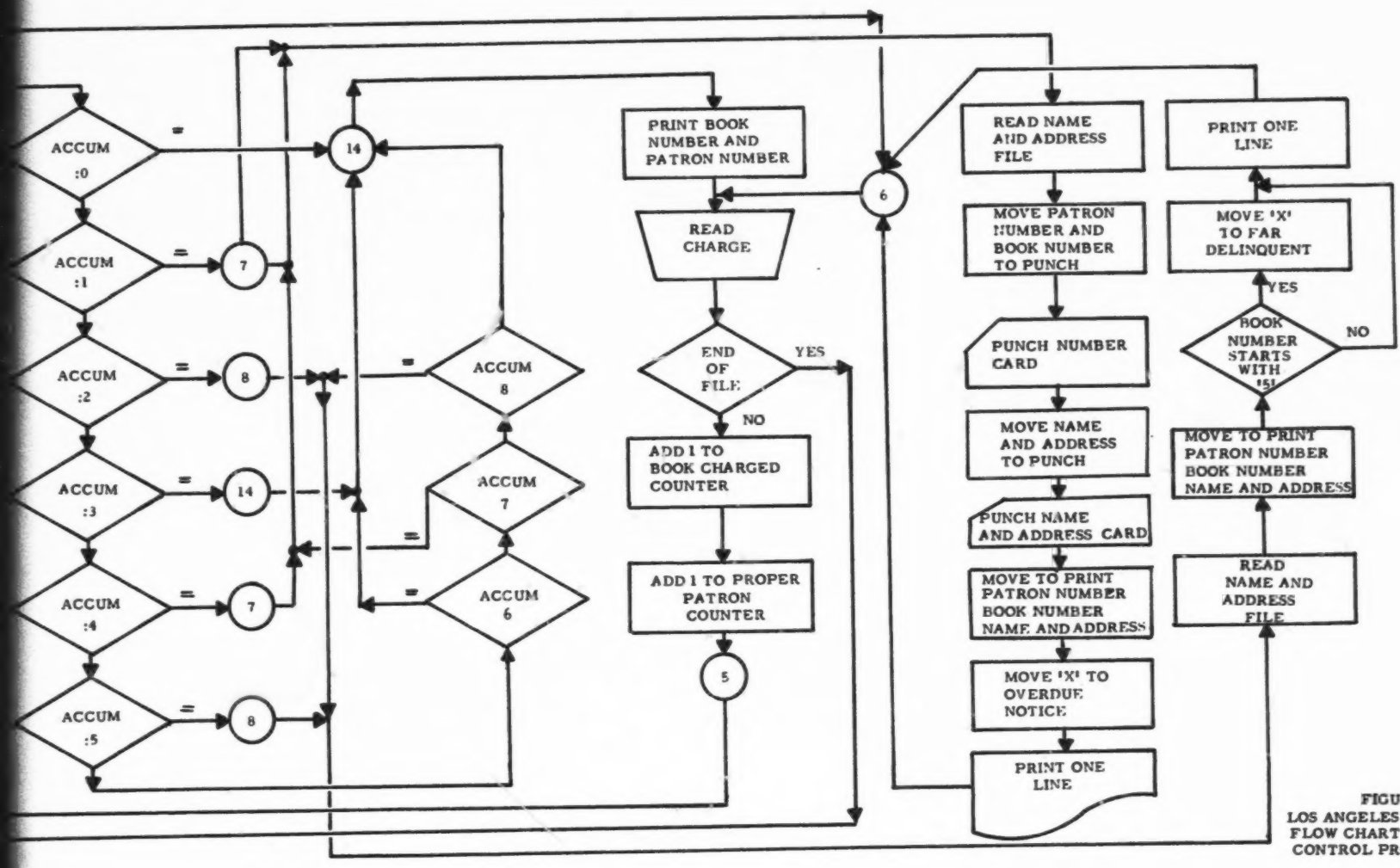


FIGURE 12
LOS ANGELES COUNTY LIBRARY
FLOW CHART - CIRCULATION
CONTROL PROGRAM

applicable, and preliminary totals are subtracted from fund accounts. Lists are then sent to vendors or the purchasing office.

The County Library has three basic procedures for processing and payment of book orders. These are:

1. Placement of a subpurchase order with a vendor with funds remaining in a blanket purchase contract. In this connection it should be noted that the County issues a number of large blanket purchase contracts to vendors who agree to supply works of specified publishers at specified discounts.

2. Placement of a subpurchase order with any vendor when costs do not exceed $25.00.

3. Placement of a requisition with the County Purchasing Office for normal bid procedures.

If items are not received in a reasonable amount of time, they are followed up with correspondence and/or cancellations. When items are received, they are unpacked, checked against invoices and made ready for further processing. Invoices are sent to the accounting office for payment. When discrepancies occur between shipment, invoices and orders, exception routines which result in correspondence and possibly delay of payment occur. After processing (cataloging, etc.) books are

shipped to regions, where order form copies are verified. Books are then distributed to local libraries in the region.

Procedures described do not pertain to periodicals, which are ordered once per year, or pamphlets which are not presented for order.

Acquisition System Conceptual Design

A computer based library acquisition system to meet the needs of the Los Angeles County Public Library has been conceptually designed using OCR input for data collection. This system is capable of operating on a small scale tape or disk oriented computer. The system design assumes batch processing on a weekly basis. The major elements of the system are:

1. Collection of typewritten, scannable source data concerning books to be ordered. Single copies of works received through approval plans would be treated on an exception basis unless multiple copy orders were anticipated. Periodicals and pamphlets would be handled as separate applications.

2. Conversion of source data to machine form, and printout of information required for order meetings.

3. Scanning of order forms and processing and printout of subpurchase order, purchase estimates, account trial balances, and reentry documents.

4. A vendor number list which would be used in creating input. Vendor information, maintained on direct access equipment, could be called out for addressing and discount calculation.

5. When operations such as receiving and cataloging, are completed, reentry documents would be batched for weekly reporting.

The basic input document to be utilized with the acquisitions system would be a variant of the bibliographical data form designed for cataloging. This would allow for the use of tags which are common to cataloging and acquisition and also for tags pertaining to only one of the two applications. This would make possible the development of an integrated system in which both cataloging and book records data could be partially captured in the acquisition processing.

Statistical and accounting reports would be produced as a by-product of the acquisitions system. Such reports could indicate such items as the number of titles purchased, processed, funds expended, and trial balances. Weekly reporting might include a report which identifies the status of each item.

SECTION IV

SYSTEMS DEMONSTRATION

In designing systems for this demonstration the following factors were considered.

1. The available equipment at the Los Angeles County Public Library. These were:
   - IBM Selectric Typewriters with the ASA font and Rabinow - CDC 915 special characters.
   - Addressograph - Multigraph two plate imprinters with embossed cards in both IBM 1428E and Farrington Self-Chek fonts.
2. The requirement for book catalog data in upper and lower case characters.
3. The availability of equipment and data centers for the demonstrations.
4. The need to simplify the programming effort.

PAGE READER DEMONSTRATION

The book catalog, book records, and registration control system were designed to be demonstrated on a page reader system. The CDC 915 was used to read the input documents and the reports were produced using an

IBM 360/30 computer with a chain printer capable of printing upper and lower case characters.

Input Preparation

The input documents were prepared by the library staff according to the specified formats. Figure 13 is a sample of the catalog input and Figure 14 is a sample of the registration control input.

The character set used for input was:

ABCDEFGHIJKLMNOPQRSTUVWXYZ

0123456789

.,:;?+=#$'/"*=()&%

The character Δ was typed in front of each alphabetic character to be output as an upper case letter.

Three identical sets of 75 documents each were typed for use in the catalog preparation demonstration. The individual sets were ordered by author, title, and subject. For the registration control system, 125 documents were prepared for use as a master file. These documents were coded as new borrower cards, and contained a patron identification code. The file was ordered by patron name. A file of 50 documents was prepared for use as the changes to the master file. These changes consisted of such items as change of address, card removal, card loss, change of name and patron type code. Only one change per entry was indicated, although in an actual system several items might change.

```
0   795.4
1   ΔABBOTT, ΔROBERT
2   ΔABBOTT'S NEW CARD GAMES.
4   ΔSTEIN, 1963.
7   ΔNINE GAMES INVENTED BY AUTHOR.
8   ΔCΔAΔRΔDΔS
K   $4.95
L   ΔLΔAΔC0 10
```

Figure 13 - Catalog Production Input

```
1    3456789218
2    AMAN, IMA JEAN
3
4    68941 2D AVENUE
5
6    CATHEDRAL CITY, CALIFORNIA 90123
7    3283809
8    1100000
```

Figure 14 - Registration Control Input

Page Reading Operation

The input documents were read twice; once to check out the system and once to demonstrate the operation to library personnel. The system checkout was performed using the optical scanner at Control Data Corporation, and the demonstration performed at the United California Bank Data Processing Center. Five tapes were produced during each operation.

- Registration Master Tape
- Registration Transaction Tape
- Author Sequence Tape
- Title Sequence Tape
- Subject Sequence Tape

Table V, on the next page, presents the breakdown of the page reader results. A total of 424 documents was read, containing 47,559 characters. There was a total of 101 reading errors when the data was read at CDC, but only 35 when the data was read at U CB. This large difference is attributed to more accurate adjustment of the reader at U CB.

As is shown in the table, the majority of the errors were in reading the number 0 and the letter B. Investigation revealed that one typewriter was slightly out of adjustment and the top of the zero and the center of the letter B were not printing properly.

Removing the errors connected with these letters resulted in significant improvement in performance. Only one other error occurred in the

TABLE V - PAGE READER DEMONSTRATION ERROR SUMMARY

Document Type	Number Document	Number Characters	Total Error		Zero's Read as U's		B's Read as Zero's		Other Errors B & Zero's		Errors Not B & Zero	
			CDC	UCB	CDC	UCB	CDC	UCB	CDC	UCB	CDC	UCB
Title Sequence	83	10,583	31	10	13	9	2	0	8	1	8	0
Subject Sequence	83	10,586	28	10	14	9	1	1	7	0	6	0
Author Sequence	83	10,581	30	13	16	9	0	0	12	4	2	0
Registration Masters	125	10,928	3(1)	0(2)	1	0	1	0	0	0	1	0
Registration Transaction	50	4,881	9	2	0	0	3	1	0	0	6	1
TOTALS	424	47,559	101	35	44	27	7	2	27	5	23	1

(1) 3 documents not read due to apparent OCR malfunction

(2) 2 documents not read due to apparent OCR malfunction

UCB reader, and 23 when read at CDC for error rates per characters read of 0.002% and 0.05% respectively.

Report Production

The tapes produced by the page reader were reformatted. This was done for ease of programming and consisted of writing one record per document, instead of the eight per document on the original tapes. The necessity for output in upper and lower case characters made it necessary to translate the data on the book catalog tapes. This translation was done by checking for the character Δ before each character in a record. For those characters not preceded by a Δ the corresponding lower case bit pattern was obtained and written on the tape.

The report production demonstration occurred on 25 January 1968 at the IBM Data Center using an IBM 360/30 computer.

The Library Registration Report, Figure 15, was produced using the registration master and the registration transaction tapes.

The author sequence tape was used to produce both the Book Catalog by Author Report, Figure 16, and the Card Printing Report, Figure 17, a & b.

The title sequence tape was used to produce the Book Catalog by Title Report, Figure 18.

Figure 19 is the Book Catalog by Subject Report which was produced using the subject sequence tape.

LOS ANGELES COUNTY LIBRARY REGISTRATION REPORT

NUMBER	PATRON NAME	ADDRESS	CITY	PHONE	CARD	COMMENT
1234567	AAROE, OSCAR	766 N. CALLE MARCUS	LOS ANGELES, CALIFORNIA 90056	3272833	ISSUE-LOST CARD	
9999023	ALBRIGHT, HENRY E.	31137 SAN LUIS REY DRIVE	LOS ANGELES, CALIFORNIA 90043	3253089	ISSUE NEW CARD	
3456789	AMAN, IMA JEAN	68941 2D A_ENUE	CATHEDRAL CITY, CALIFORNIA 90123	3283809	ISSUE-NAME CHANGE	
1234568	ANDERSON, ROBERT M.	2125 CALLE LILETA	HAWTHORNE, CALIFORNIA 90250	3255406	ISSUE-LOST CARD	
3456780	ARREOLA, BENNY A.	279 N. CALLE ENCELIA	PALMDALE, CALIFORNIA 98175	3245115	ISSUE-LOST CARD	
1234569	BABB, MILDRED A.	66781 VISTA PLACE	LANCASTER, CALIFORNIA 94327	3294463	ISSUE-NAME CHANGE	
3145678	BECKER, GLORIA	1905 MCCALLUM WAY	LANCASTER, CALIFORNIA 94327	3243229	ISSUE-NAME CHANGE	
1234560	BEEVER, SYDNEY	68159 TERRACE ROAD	CATHEDRAL CITY, CALIFORNIA 90123	3282054	ISSUE-RENEW CARD	
3245678	BENNETT, JOHN S.	1164 N. ROSE AVENUE	INGLEWOOD, CALIFORNIA 90251	3249431	ISSUE-RENEW CARD	
1234561	BIRNKRANT, MICHAEL C.	853 N. PATENCIO ROAD	COVINA, CALIFORNIA 98725	3256480	ISSUE-RENEW CARD	
3045678	BORBETT, ATHEARN	1171 N. PLAYA DEL REY STREET	ALHAMBRA, CALIFORNIA 90111	3256683	ISSUE-NEW CARD	
3334567	BREEDING, HARRY J.	36956 MELROSE DRIVE	ALHAMBRA, CALIFORNIA 90111	3281518	DELINQUENT	
1234562	BURNS, MABLE	34532 MARGUERITE DRIVE	CATHEDRAL CITY, CALIFORNIA 90123	3281262	ISSUE-NAME CHANGE	
1234563	BYRD, SOPHIA JANE	21966 SNOW VIEW DRIVE	LOS ANGELES, CALIFOR_IA 90021	3255063	ISSUE-ADDRESS CHANGE	
1234564	CADDLE, ETHELYN	65935 OUE_A VISTA AVENUE	LANCASTER, CALIFORNIA 93534	3295552	ISSUE-ADDRESS CHANGE	
1234565	CANTRELL, EDNA	66462 FLORA AVE_UE	LITTLEROCK, CALIFORNIA 93543	3295537	ISSUE-ADDRESS CHANGE	
3345678	CARSON, DEWEY L.	66856 DESERT VIEW AVENUE	SAN FERNANDO, CALIFORNIA 91340	3295327	ISSUE-ADDRESS CHANGE	
3554567	CHICK, JUNE HOGAN	2251 AVENIDA CABALLEROS	QUARTZ HILL, CALIFORNIA 93534	3244671	ISSUE-ADDRESS CHANGE	
1234566	CLARK, JOHN PORTER	1200 PASEO EL MIRADOR	QUARTZ HILL, CALIFORNIA 93534	3249425	ISSUE-ADDRESS CHANGE	
1234511	COMSTOCK, HERBERT	67160 SAN GABRIEL CIRCLE	PALMDALE, CALIFORNIA 93550	3272341	ISSUE-ADDRESS CHANGE	
3654567	COPELAND, JULIANNA H.	2842 STARR ROAD	LITTLEROCK, CALIFORNIA 93534	3253795	ISSUE-ADDRESS CHANGE	
1234512	CROCKETT, MERRILL	456 W. ARE_AS ROAD	PALMDALE, CALIFORNIA 93550	3253093	ISSUE-ADDRESS CHANGE	
3776548	DEARING, LESTER W.	70260 HIGHWAY 111	HAWTHORNE, CALIFORNIA 90250	3283250	ISSUE-ADDRESS CHANGE	
1234513	DEDRICK, LULU B.	3865 E. PALM CANYON DRIVE	LENNOX, CALIFORNIA 90304	3283623	ISSUE-NEW CARD	
3887654	DERRY, HENRIETTA C.	66822 DESERT VIEW AVENUE	LENNOX, CALIFORNIA 90304	3282596	ISSUE-NEW CARD	

Figure 15 (a)

PATRON STATISTICAL REPORT - LIBRARY-101

1	ADULT-FEMALE	SNS*	15
2	----------	MNS	13
3	----------	SS	10
4	----------	MS	10
5	ADULT---MALE	SNS	14
6	----------	MNS	17
7	----------	SS	16
8	----------	MS	10
9	JUVENILE--	MALE	14
0	----------FEMALE		16

TOTAL PATRONS 135

*Patron Identification Code:

SNS - Single, Non Student

MNS - Married, Non Student

SS - Single, Student

MS - Married, Student

Figure 15 (b)

LOS ANGELES COUNTY LIBRARY
BOOK CATALOG PRODUCTION----AUTHOR SEQUENCE

Aaron, David
Child's play.
Harper, 1965.

641.6 Aaron, Jan
Chicken cookbook.
Doubleday, 1963.

362.7 Aaronson, Sammy
As high as my heart.
Coward, 1957.

581.937 Abbe, Elfriede Martha
Plants of Virgil's Georgics.
Cornell Univ. Pr., 1965.

598.8 Abbe, George
Larks.
Regnery, 1965.

R598.297 Abbott, Charles Conrad
Birds about us.
Lippincott, 1894.

513.8 Abbott, Edwin Abbott
Flatland.
Blackwell, 1950.

Figure 16

Aaron, David
Child's play.
Harper, 1965. $3.95

1		6							
2		7							
3		8							
4		9							
5		10							

LOCATION CARD

Aaron, David
Child's play.
Harper, 1965. $3.95

1		12		23		34	
2		13		24		35	
3		14		25		36	
4		15		26		37	
5		16		27		38	
6		17		28		39	
7		18		29		40	
8		19		30		41	
9		20		31		42	
10		21		2		43	
11		22		3		44	

MASTER SHELF LIST CARD

Aaron, David
Child's play.
Harper, 1965. $3.95

SPINE LABELS

Figure 17(a)

Aaron, David
Child's play.
Harper, 1965. $3.95

CATALOG OR SHELF LIST CARD

Aaron, David
Child's play.
Harper, 1965. $3.95

BOOK CARD

Aaron, David
Child's play.
Harper, 1965.- $3.95

POCKET LABEL

Figure 17(b)

LOS ANGELES COUNTY LIBRARY
BOOK CATALOG PRODUCTION----TITLE SEQUENCE

Abbott's new card games.
Abbott, Robert

636 Animal science and industry.
Acker, Duane

362.7 As high as my heart.
Aaronson, Sammy

574 Biology.
Alexander, Gordon

R598.297 Birds about us.
Abbott, Charles Conrad

9u4 Broadsides and boarders.
Albert, Marvin H.

529.5 Calendar for the modern age.
Achelis, Elisabeth

S784.4 Canciones populares Chilenas.
Acevedo Hernandez, Antonio

641.6 Chicken cookbook.
Aaron, Jan

Figure 18

LOS ANGELES COUNTY LIBRARY
BOOK CATALOG PRODUCTION----SUBJECT SEQUENCE

AARONSON, SAMMY
362.7 Aaronson, Sammy
As high as my heart.
Coward, 1957.

ABBOTT, GEORGE
92 A1314 Abbott, George
Mister Abbott.
Random House, 1963.

ANIMAL INDUSTRY
636 Acker, Duane
Animal science and industry.
Prentice, 1963.

ANIMALS
59u Aistrop, Jack Bentley
Fun at the zoo.
Roy, 1951.

ARCHAEOLOGY
571 Aitken, Martin Jim
Physics and archaeology.
Interscience Pub., 1961.

BAHAISM
299.55 'Abd ul Baha ibn Baha Ullah
Some answered questions.
Baha'i, 1964.

Figure 19

DOCUMENT READER DEMONSTRATION

The circulation control system was demonstrated using a document reader. The reader was an IBM 1282 document reader.

Input Preparation

The two plate imprinter was used to prepare approximately 1000 charge and 750 discharge documents. These documents were two part, interleaved tab stock cards. The documents normally used by the library were on paper which could not be processed by the IBM 1282. The documents used were especially purchased for the demonstration. Each contained a patron name, patron number, and the book number. The discharge documents were a subset of the charge documents.

If a document reader with mirror-image reading capability had been used one part of the form would be the charge document and the other part would be the discharge document.

Document Reader Operation

The document reader demonstration occurred at the Richfield Data Processing Center in Los Angeles. The reader read the card and punched the information on the same card.

The charge file of 1044 documents was first processed with 253 documents rejected. Of the 783 documents in the discharge file, 209 were

rejected. To understand this rejection percentage, 24.2% and 26.7% respectively, an explanation of the problems encountered in the data preparation stage is necessary.

- The two plate imprinter was not adjusted to compensate for the difference in the thickness and length of the library form and the form used for the demonstration.
- The placement of the data on the forms was at the extreme limits of the scanning field of the reader.
- One box of the forms was marked by smudging from greasy carbon and careless handling.
- The document reader was not adjusted for this type of application.

In view of these results, investigations were conducted by DMJM, in order to determine what could be expected from this type of document reader under normal circumstances. It was found that in oil credit card applications, a rejection rate of 15 to 25% is acceptable. It is DMJM's conclusion, however, that with proper care in data preparation, with a fully adjusted imprinter, with adequate forms and carbon paper, and using a document reader adjusted for correct print intensity, the library could have achieved a reject rate of less than 5% in this type of application. The "charge" part of the form should present no problems since it will remain in file until processing. The "discharge" part of the form circulated with the book could, upon return for processing, be smudged or marked. The quality of the carbon used should be of particular concern.

Report Production

The Book Circulation Report was processed at DMJM on the UNIVAC 9200 with a 1001 card controller on line. The documents rejected during the reading stage were keypunched, sorted, and merged into the charge and discharge files.

Even though the punched cards were used to produce the report, the magnetic tape concept was respected. The 1001 card controller primary was loaded with the "discharge" file, and the secondary with the "charge" file. In the reader was loaded a "name and address" file representing a magnetic tape borrower Master File which would be searched for report printing and "overdue notices" preparation.

In an actual situation, overdues would be sorted by borrower number after determining overdue status. This would allow all books an individual had overdue to be listed on a single overdue notice after borrower name and address were retrieved from the borrower Master File.

The Circulation Report is presented in Figure 20a. An additional output from the Circulation Report is a statistical report on the transaction. This report is presented in Figure 20b.

Figure 21 presents the overdue notices as prepared by the cards output from the Circulation Report program.

O U T P U T R E P O R T

IN CIRCULATION BOOK	BORROWER	BORROWER NAME	ADDRESS		OVERDUE NOTICE	FAR DELQT
1234567	2000001					
1234567	5526098	OAKS, RUTH J.	410 PASEO EL MIRADOR	LAVERNE, CALIF 91750		X
1234569	2000001	CLARK, JOHN PORTER	1200 PASEO EL MIRADOR	QUARTZ HILL, CALIF 93534	X	
1234569	3765408	FAIRCHILD, ADELINE	420 S. BELARDO ROAD	TORRANCE, CALIF 90501		
1234571	7332175	WEGENER, FRIEDA	390 LAUREL CIRCLE	PARAMOUNT, CALIF 91156		
1234571	9337187	MCGLANARY, PENNY	511 N. VIA MIRALESTE AVEN	ARCADIA, CALIF 91006	X	
1234572	3876519					
1234573	5543286					
1234574	2000002	ELLIOTT, JACKSON S.	321 VIA DEL SOL	HAWTHORNE, CALIF 90250		
1234574	9337187					
1234576	9660432	KORODY, PAUL A.	591 N. PATENCIO ROAD	WHITTIER, CALIF 98111		
1234577	9182654					
1234578	7102874					
1234578	5006854	UNTERMYER, ALVIN	412 W. TAHQUITZ DRIVE	INGLEWOOD, CALIF 90302	X	
1234578	9102874	QUIROZ, RICHARD	33305 RANCHO VISTA DRIVE	HAWTHORNE, CALIF 90250	X	
1234579	9102874					
1234579	9102787	YERKES, ESTELLE F.	633 CHINO CANYON DRIVE	INGLEWOOD, CALIF 90302	X	
1234581	1000002					
1234580	9102874	PALMER, BARBARA BRADLEY	733 CAMINO SUR	SAN FERNANDO, CALIF 91340	X	
1234580	9952704	TURNER, BETTY F.	443 N. PALM CANYON DRIVE	HUNTINGTON PARK, CALIF 90256	X	
1234581	7210953					
1234581	5007215	WHITE, MARK S.	593 CALLE ABRONELA	LAWNDALE, CALIF 90260		X
1234583	2000001	WAGNER, ALBERT C.	404 CAMINO DEL SUR	INGLEWOOD, CALIF 90302	X	
1234583	9273666					
1234583	9526098	MARSHALL, PEGGY	634 N. CALLE DE ROLPH	CULVER CITY, CALIF 90330		
1234584	7210953					

Figure 20 a.

STATISTICAL REPORT

BOOKS CHARGED.....	1045
BOOKS RETURNED.....	785
IN CIRCULATION.....	260

CIRCULATION BY DEWEY NO.

000'S.....	24
100'S.....	90
200'S.....	90
300'S.....	166
400'S.....	25
500'S.....	100
600'S.....	150
700'S.....	124
800'S.....	126
900'S.....	150

Figure 20 b

LOS ANGELES COUNTY LIBRARY 01-25-68

DEAR FRIEND,

OUR RECORDS INDICATE THAT THE BOOK IDENTIFIED BELOW IS OVERDUE. IF YOU HAVE RETURNED THIS ITEM IN THE LAST FEW DAYS, PLEASE DISREGARD THIS NOTICE

BORROWER 2000001 — BOOK 1234569

CLARK, JOHN PORTER
1200 PASEO EL MIRADOR
QUARTZ HILL, CALIF 93534

LOS ANGELES COUNTY LIBRARY 01-25-68

DEAR FRIEND,

OUR RECORDS INDICATE THAT THE BOOK IDENTIFIED BELOW IS OVERDUE. IF YOU HAVE RETURNED THIS ITEM IN THE LAST FEW DAYS, PLEASE DISREGARD THIS NOTICE

BORROWER 9337187 — BOOK 1234571

MCGLANARY, PENNY
511 N. VIA MIRALESTE AVEN
ARCADIA, CALIF 91006

LOS ANGELES COUNTY LIBRARY 01-25-68

DEAR FRIEND,

OUR RECORDS INDICATE THAT THE BOOK IDENTIFIED BELOW IS OVERDUE. IF YOU HAVE RETURNED THIS ITEM IN THE LAST FEW DAYS, PLEASE DISREGARD THIS NOTICE

BORROWER 5006854 — BOOK 1234578

UNTERMYER, ALVIN
412 W. TAHQUITZ DRIVE
INGLEWOOD, CALIF 90302

LOS ANGELES COUNTY LIBRARY 01-25-68

DEAR FRIEND,

OUR RECORDS INDICATE THAT THE BOOK IDENTIFIED BELOW IS OVERDUE. IF YOU HAVE RETURNED THIS ITEM IN THE LAST FEW DAYS, PLEASE DISREGARD THIS NOTICE

BORROWER 9102874 — BOOK 1234580

BUIROZ, RICHARD
33105 RANCHO VISTA DRIVE
HAWTHORNE, CALIF 90250

LOS ANGELES COUNTY LIBRARY 01-25-68

DEAR FRIEND,

OUR RECORDS INDICATE THAT THE BOOK IDENTIFIED BELOW IS OVERDUE. IF YOU HAVE RETURNED THIS ITEM IN THE LAST FEW DAYS, PLEASE DISREGARD THIS NOTICE

BORROWER 9952704 — BOOK 1234580

YERKES, ESTELLE F.
633 CHINO CANYON DRIVE
INGLEWOOD, CALIF 90302

LOS ANGELES COUNTY LIBRARY 01-25-68

DEAR FRIEND,

OUR RECORDS INDICATE THAT THE BOOK IDENTIFIED BELOW IS OVERDUE. IF YOU HAVE RETURNED THIS ITEM IN THE LAST FEW DAYS, PLEASE DISREGARD THIS NOTICE

BORROWER 2000001 — BOOK 1234583

TURNER, BETTY F.
441 N. PALM CANYON DRIVE
HUNTINGTON PARK, CALIF 90256

Figure 21 - Overdue Notices

902

PLEASE PRINT OR CHECK BELOW

MR. ☐
MRS. ☐
MISS ☐ LAST NAME

FIRST NAME | MIDDLE NAME OR INITIAL | AGE

MAILING ADDRESS

CITY | ZIP CODE | TELEPHONE

☐ NEW CARD ☐ RENEWAL ☐ ADDRESS CHANGE
☐ LOST CARD ☐ NAME CHANGE ☐ CHECK IF YOU ARE A STUDENT

I AM WILLING TO ASSUME THE RESPONSIBILITY FOR LIBRARY MATERIALS BORROWED FROM THE LOS ANGELES COUNTY PUBLIC LIBRARY

LEGAL SIGNATURE

PARENT OR GUARDIAN

DO NOT WRITE BEYOND THIS LINE

LOS ANGELES COUNTY PUBLIC LIBRARY
PATRON REGISTRATION

NUMBER | LIB.

FOR LIBRARY USE ONLY

NAME

PRIOR NAME | AGE

ADDRESS (STREET)

PRIOR STREET ADDRESS

CITY | ZIP | TELEPHONE | CD

☐ NEW CARD | ☐ LOST CARD | ☐ RE-NEW | ☐ NAME CH. | ☐ ADDR. CH. | ☐ DEL. | ☐ REN.

LIBRARY COMMENTS:

LACO-CLR 902

Figure 22
Patron Registration Form

901

LOS ANGELES COUNTY PUBLIC LIBRARY
CIRCULATION TRANSACTION

DATE DUE

BORROWER

BOOK

LACO-CLR 901

Figure 23
Circulation Transaction Form
(A two part form)

LOS ANGELES COUNTY PUBLIC LIBRARY
BIBLIOGRAPHICAL DATA

Figure 24
Bibliographical Data Form

SECTION V

PROJECT SUMMARY

Research and development of prototype library systems which utilize OCR input has centered around optical page readers and document readers. The state-of-the-art of both types of readers makes them currently acceptable for library input preparation. The potential uses of page readers include:

1. Patron registration
2. Ordering of materials
3. Serials control
4. Bibliographic data collection

Document Readers may be used in charging and discharging functions of a circulation system.

Demonstrations conducted during the course of the project indicate that with proper preparation and normal attention to detail, the error and rejection rates of both page and document readers can be kept well within desirable limits. Error detection systems for document readers require more care.

Page Readers

The most complex library application for page readers is bibliographic data collection when upper and lower case characters, numerous symbols and foreign alphabets are used. Scanners with proper control procedures are very accurate. It is currently possible to rent time on page readers for $30/hour or approximately $.02 per line scanned. However, equipment generally available is limited to upper case roman alphabet, numbers and a maximum of 18 symbols in a stylized font. Output from such scanners is usually IBM compatible magnetic tape.

An OCR system similar to the above would rent for approximately $3,500/ month. Depending on the speed of the system, multi-font capabilities and

other features, an OCR page reading system rents for as much as $15,000/ month. It is also currently possible to acquire a complete computer system which includes CPU (16,000 chars. storage), 4 tape drives, printer (600 lpm) and OCR page reader for approximately $5,000/month.

Document Readers

The potential for library utilization of document readers is much more limited than page readers. The major application for such readers envisions plastic embossed borrower and book plates used in conjunction with imprinters to produce optically readable imprinted documents (charges and discharges). Conversion costs for plates, the lack of completely suitable imprinters, and lack of document readers at service bureaus make it difficult for libraries to implement such a system.

Rental of a document reader capable of performing this function starts at about $3,000/month. Even with a well controlled document preparation procedure a reject rate of 2% on imprinted documents seems imminent. Such a system becomes practical with a great number of charging stations.

Advantages of Page Readers

OCR page readers have certain advantages over other off-line input methods such as keypunches, paper tape punching typewriters and magnetic tape writers. Specifically, these advantages are:

1. OCR provides the only method in which keyed output is both man and machine readable.
2. OCR provides the potential for most flexible error correcting

procedures prior to a machine run. This may include use of correction tape in typing, penciling out of incorrect or unwanted data and addition of data at the end of input documents for computerized reformatting.

3. Complete, lengthy, logical records may appear on one document as opposed to a number of cards in keypunching. This requires less controls such as card number, a repeated entry number, etc.
4. Formatting of records is more flexible. Pages of various sizes may be arranged to combine machine readable data with other information. For example, a source document which may be handwritten on in different areas by a number of persons, may then be typed on for machine reading.
5. In one-time data conversion and as the amount of data input increases, OCR becomes more economically attractive than other off-line data collection methods.

It is difficult to generalize about the differences between on-line keyboard systems as opposed to page readers. This is due to the fact that on-line keyboards are available with a very wide range of capabilities, special features, price ranges, etc.

Advantages of Document Readers

OCR document readers have advantages over 'on' or 'off-line' transactor devices. Specifically, these advantages are:

1. Low cost imprinter charging/discharging stations may be quickly located, relocated or added without difficulty. The number of stations has little effect on feasibility of system.

2. In using imprinted documents, which are man as well as machine readable, manual backup for the system is more fairly easily accomplished.

OCR Input For Libraries In The Future

OCR page readers will be used increasingly by libraries and other organizations in the future. Use will be most widespread for single font, alphameric readers, with a full set of special symbols and the capability for discrete recognition and encoding of upper and lower case alpha characters. Special readers with multi-lingual and multi-font capability will be used by large bibliographic centers and specialized agencies. Some of these machines may be available on a service basis.

To a lesser extent, document readers will be utilized by large systems of libraries. An impetus for this will be the consolidation of credit card issuing agencies and increasing dependence on an expansion of the social security number as an indentifier for the individual.

While there is much interest in on-line real-time data collection systems, it seems highly unlikely that off-line batch type data collection systems will be eliminated from the data processing scene. Cost factors, technical problems and operational impracticality of collecting certain types of data on-line necessitate the maintenance of off-line batch systems. OCR, as a result, will continue to grow in importance as a data collection tool.